THE UNIVERSITY
of EDINBURGH

Structural basis of MeCP2 interaction with NCoR/SMRT co-repressor complex

Valdeko Kruusvee (BScH)

Submitted for the degree of Doctor of Philosophy

September 2016

Wellcome Trust Centre for Cell Biology

University of Edinburgh

Acknowledgments

I would like to thank both my supervisors, Dr. Atlanta Cook and Prof. Adrian Bird for their support and patience during this PhD, as well as for their constructive criticism to help me become a better scientist. I would also like to thank Dr. Matthew Lyst, with whom I collaborated during this project and whose previous work was instrumental to this project. I would also like to extend my gratitude to my thesis committee members, Dr. Laura Spagnolo and Prof. Paul Barlow.

My deepest gratitude goes to Rett Syndrome Research Trust for funding this research, and to both Rett Syndrome Research Trust and The University of Edinburgh for providing me with this opportunity.

I am grateful to the past and present members of the Cook lab for their continued support and help. I would like to thank both Ceitidh Taylor, our Honours student, and Ciaran Daly, our summer student, who helped with the purification and biophysical assays of TBL1XR1 mutants and MeCP2 minigene protein, respectively.

I would like to express my gratitude to the members of the Edinburgh Protein Purification Facility and Biophysical Characterisation Laboratory for the help and training they provided over the years. Finally, special thanks to all the people on the third floor of Swann building for their help and good chat.

Declaration

The work presented in this thesis was conducted solely by the author. Any contributions by other people have been clearly indicated as such. The work has not been submitted for any other degree or professional qualification.

Signed,

Valdeko Kruusvee

Table of Contents

Table of Figures

Table of Tables

a. Lay summary

Dysregulation of brain function can give rise to a variety of disorders. One of these disorders is called Rett syndrome, which is one of the most common causes of mental retardation in females. It gives rise to symptoms that are similar to many autism spectrum disorders and it is caused by mutations in a protein called MeCP2. People with mutated MeCP2 protein have developmental problems in the brain. After a period of normal development, they regress with symptoms such as lack of speech, loss of motor skills, hand wringing and breath holding. Females with Rett syndrome can live up to 50 years and more, while males usually die before the age of 2.

However, it has been shown that the developmental problems caused by MeCP2 mutations are reversible in mice, and potentially in humans as well. What MeCP2 does in cells is not completely understood. Recently, it has been found that MeCP2 has a partner protein called TBL1XR1 and that some MeCP2 mutations prevent it from binding to its partner in cells. This project aimed to understand why this interaction is lost when MeCP2 is mutated. I looked at the three-dimensional structure of the two proteins when they interact, using a method known as X-ray crystallography. From this a detailed understanding of how normal MeCP2 interacts with TBL1XR1 was gained. I also measured how strongly the normal MeCP2 and TBL1XR1 proteins interact, and how well the mutated MeCP2 protein interacts with TBL1XR1. Through this I show that the mutations in the MeCP2 stop it from binding to TBL1XR1. Finally, I show that mutated TBL1XR1 is equally able to stop the interaction with normal MeCP2.

Through this a thorough understanding of how these known MeCP2 mutations stop it from interacting with TBL1XR1 was established. This knowledge can be used to develop drugs that combat this disease and potentially treat it.

b. Abstract

Rett syndrome (RTT) is an X-linked neurological disorder primarily caused by mutations in the *MECP2* gene. The majority of RTT mutations disrupt the interaction of MeCP2 with DNA or TBL1X/TBL1XR1, which forms the scaffold of NCoR/SMRT co-repressor complex. Patients with RTT show no signs of neuronal death, although they have abnormal neuronal morphology, indicating that it is a neurodevelopmental rather than a neurodegenerative disease. It has been shown that reactivation of silenced MeCP2 in mice rescues the RTT phenotype, which implies that the disease is treatable. The RTT mutations in MeCP2 cluster to two regions - the methyl-CpG-binding domain (MBD) and NCoR/SMRT Interaction Domain (NID). While the interaction between MBD and DNA has been biochemically and structurally characterised, there are no structural data about the interaction between MeCP2 NID and TBL1XR1.

The aim of this work was to understand how mutations in the NID cause RTT by characterising the interaction between MeCP2 and TBL1XR1. I have solved the structure of MeCP2 NID bound to TBL1XR1 WD40 domain. I show that a small region of the MeCP2 NID makes extensive contacts with TBL1XR1, and that these contacts are mediated primarily by MeCP2 residues known to be mutated in RTT. I also measured the affinities between TBL1XR1 and MeCP2-derived peptides using fluorescence anisotropy and surface plasmon resonance assays. I determined the affinity between MeCP2 NID peptide and TBL1XR1 to be around 10-20 μM, and show that mutations in either MeCP2 or TBL1XR1 can abolish this interaction.

Taken together, these data strongly suggest that the abolition of the interaction between MeCP2 NID and TBL1XR1 WD40 domain is sufficient to cause RTT. This knowledge can help with the rational design of small drug-like molecules that might be able to mediate the interaction between mutated MeCP2 and TBL1XR1, potentially helping to treat the disease.

c. Abbreviations

5caC - 5-carboxylcytosine

5fC - 5-formylcytosine

5hmC - 5 hydroxymethylcytosine

5mC - 5-methylcytosine

ADHD - Attention-deficit hyperactive disorder

AmSO4 - ammonium sulphate

Bacmid - Bacterial Artificial Chromosome

BDNF - Brain-derived neurotrophic factor

BER - Base Excision Repair

BSA - bovine serum albumin

CBP - CREB-binding Protein

CpG - CG dinucleotide

DAD - Deacetylase Activation Domain

DMSO - Dimethylsulfoxide

DNA - Deoxyribonucleic acid

DNMT3a/b - De Novo Methyltransferase 3a/b

dNTP - deoxynucleoside triphosphate

DTT - Dithiothreitol

EDTA - Ethylenediaminetetraacetic acid

EMSA - Electrophoretic Mobility Shift Assay

FA - Fluorescence Anisotropy

GDD - Global Development Delay

GPS2 - G-protein Pathway Suppressor Protein 2

HAT - Histone Acetyl Transferase

HDAC - Histone Deacetylase

IMAC - Immobilised Metal Affinity Chromatography

IP4 - Inositol tetraphosphate

IPTG - Isopropyl beta-D-1-thiogalactopyranoside

K_D - Dissociation constant

kDa - kilodalton

LB - Lysogeny Broth

LIC - Ligation-independent Cloning

LisH - LIS1 Homology

MALDI TOF-TOF - Matrix-assisted Laser Desorption Ionisation Time-Of-Flight Time-Of-Flight

MBD - Methyl-CpG-binding Domain

mCpG - methylated CG dinucleotide

MeCP2 - Methyl-CpG-binding Protein 2

NuRD - Nucleosome Remodelling Deacetylase

MWCO - Molecular Weight Cut-off

NCoR - Nuclear receptor co-repressor

NES - Nuclear Export Signal

NID - NCoR/SMRT Interaction Domain

NLS - nuclear localisation signal

OD - Optical Density

PCR - Polymerase Chain Reaction

PDB - Protein Data Bank

pI - Isoelectric point

RING - Really Interesting New Gene

RTT - Rett syndrome

RU - response units

S/N - Supernatant

SDS-PAGE - Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC - Size-exclusion Chromatography

SMRT - Silencing mediator for retinoid or thyroid-hormone receptors

SPR - Surface Plasmon Resonance

TBL1X - Transducin-beta-like protein 1, X-linked

TBL1XR1 - Transducin-beta-like 1X-related protein 1

TBL1Y - Transducin-beta-like protein 1, Y-linked

TDA - Thermal Denaturation Assay

TET - Ten Eleven Translocase

Tm - Melting temperature

1 Introduction

1.1 Rett Syndrome

Rett syndrome (RTT) is an X-linked neurological disorder characterised by initial normal development for about 6-9 months, followed by gradual loss of cognitive and psychomotor skills. Symptoms of RTT include seizures, loss of purposeful hand movement, breath holding and loss of speech, among others (Samaco and Neul, 2011). RTT is a monogenic disorder caused by mutations in the X-linked *MECP2* gene (Amir et al., 1999). Familial inheritance of RTT is rare as the affected females have lower reproduction capability (Dragich et al., 2000), with the majority of mutations being sporadic and occurring on the paternal X-chromosome (Trappe et al., 2001). As males inherit their X-chromosome from their mothers, this leads to highly skewed male:female ratio of RTT patients. Additionally, the loss of the only functioning copy of MeCP2 in males causes severe neonatal encephalopathy and leads to early lethality. Together, this means that the disorder primarily affects females (Villard, 2007). Despite this, in rare cases RTT can occur in males with abnormal karyotype, e.g. Klinefelter syndrome (Villard et al., 2000) The severity of RTT in females is variable due to skewed X-inactivation and, in rare cases, this can lead to complete absence of the RTT phenotype, which contributes to familial cases of RTT (Figure 1.1.1) (Villard et al., 2000).

Individuals with Rett syndrome show no signs of neuronal death, although they do display altered neuronal morphology, including smaller soma and a reduced number of dendritic spines, which is consistent with a neurodevelopmental disorder (Marchetto et al., 2010, Armstrong et al., 1995). However, it has been shown that reactivation of MeCP2 expression in mice containing a silenced endogenous MeCP2 gene leads to a reversal of RTT symptoms. This suggests that the change in the morphology of neurons caused by the absence of MeCP2 during development is not irreversible (Guy et al., 2007). This implies that it is theoretically possible to treat Rett syndrome in humans.

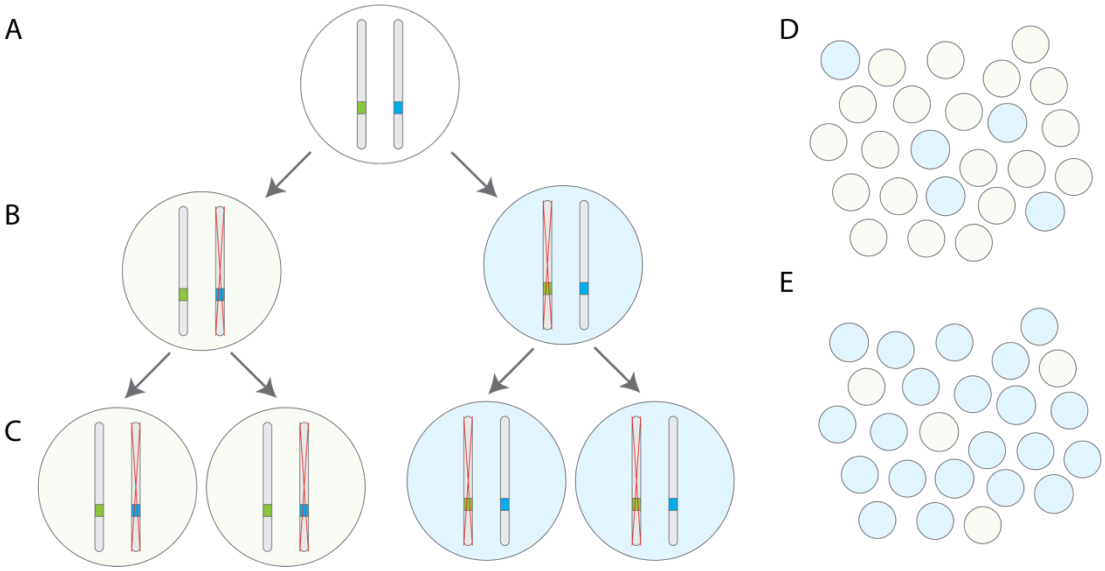

Figure 1.1.1. X-inactivation in females. A) A cell with two X-chromosomes and no established X-inactivation, one with a functioning copy (green), and one with a mutant copy (blue). B) At some point in early zygote, one of the X-chromosomes gets randomly irreversibly inactivated, resulting in one cell being wild-type, and one cell being mutant. C) All future progeny of these cells have the same copy of the X-chromosome inactivated, resulting in stable inheritance. D and E) Skewing of X-inactivation can cause variable severity of RTT, from very mild/asymptomatic (D), to severe RTT (E).

While mutations in MeCP2 can cause RTT, duplication of MeCP2 leads to a disease known as MeCP2 overexpression syndrome. MeCP2 overexpression syndrome primarily affects males and causes severe intellectual disability with autistic features, similar to RTT (Van Esch, 2012, Van Esch et al., 2005). Recently it has been shown that MeCP2 duplication is only detrimental when wild-type protein is overexpressed, as co-expressing one copy of mutated MeCP2 with one copy of wild-type produced no phenotype (Heckman et al., 2014).

There are reported cases of patients with RTT who do not have mutations in MeCP2 (Temudo et al., 2011), and likewise patients with RTT-causing MeCP2 mutations who do not have RTT (Suter et al., 2014). In the first case, it is possible that other disorders have been classified as RTT, or that RTT is not monogenic, although is currently classified as such. In the second case, some unknown factor could have caused phenotypic variation in the presence of well-characterised RTT mutations. Other mutations in MeCP2 that do not cause RTT have been proposed to cause X-linked mental retardation (Couvert et al., 2001), and potentially global development delay (GDD) and attention-deficit hyperactive disorder (ADHD) (Suter et al., 2014).

RTT is a clinical diagnosis that does not require mutations in *MECP2*. Patients who do not fulfil all the criteria for RTT are described as having atypical, or variant, RTT (Neul et al., 2010).

However, approximately 95% of patients with RTT have mutations in the MeCP2 gene. Additionally, just eight MeCP2 mutations, which disrupt either one of the two well-characterised domains described later in this thesis, make up over 60% of RTT cases, making the correlation between MeCP2 mutations and RTT very strong (Neul et al., 2008).

One view is that MeCP2 is involved in a multitude of interactions with various proteins and the various phenotypes, both RTT and others not classified as such, arise from defects in these specific interactions. The opposing view is that MeCP2 is involved in just RTT, and that the various mutations disrupt two known functions of this protein. Understanding what the different regions of MeCP2 do can provide information about how these regions contribute to the phenotypes observed.

1.2 Overview of gene transcription regulation

MeCP2 was first identified as a protein that binds to methylated CpG dinucleotides in DNA (Lewis et al., 1992). This implies that it may be involved in transcriptional control as levels of gene transcription are thought to be influenced by DNA methylation and histone modifications, referred to as the epigenetic code (Turner, 2007). Control over gene transcription is essential for the development and survival of an organism. A predominant DNA modification that is heritable between cell divisions is 5-methylation of cytosines (5mC) in CG dinucleotide context. This modification is added by *de novo* methyltransferases DNMT3A and DNMT3B (Okano et al., 1999), and maintained by DNMT1 (Hermann et al., 2004). In most metazoans DNA methylation occurs in gene bodies while in vertebrates the genome is globally methylated (Suzuki and Bird, 2008). Most of the genome is methylated at low density (one mCpG per ~100bp) due to the rarity of CpG itself (Illingworth et al., 2008). This is a consequence of the mutability of mCpG, where spontaneous deamination of 5mC converts it into thymidine, and so CpG dinucleotides are selected against in evolution and are underrepresented in the genome.

However, certain regions called CpG islands (CGIs) contain a large number of unmethylated CpGs (Illingworth et al., 2008). Roughly 50% of CGIs are associated with annotated transcription start sites, with the rest (called orphan CGIs) occurring in intra- and intergenic regions. Some of these orphan CGIs have been found to represent novel transcription start sites (Smith and Meissner, 2013, Deaton and Bird, 2011). Methylation of CGIs is thought to be required to reinforce silencing, as most CGIs that are methylated are already transcriptionally silent prior to methylation. However, unmethylated CGIs can also be

transcriptionally inactive (Jaenisch and Bird, 2003, Weber et al., 2007). CGIs are also linked to histone modifications, as the presence of CGIs is correlated with H3K4me3, and H3/H4 acetylation, both of which are associated with active transcription (Deaton and Bird, 2011).

DNA methylation is read by proteins belonging to the methyl-CpG-binding domain (MBD) and Kaiso protein families. Members of the MBD family are MBD1, MBD2, MBD3, MBD4 and MeCP2, which all possess homologous methyl-CpG-binding domains (MBDs). DNA methylation has a multifaceted role. Methylation of DNA leads to transcriptional repression through either direct blocking of transcription factors, or through the exclusion of transcription factors by the methyl-DNA binding proteins. The latter can, in turn, lead to recruitment of proteins that modify either chromatin or histones. For example, the MBD family proteins interact with proteins involved in transcriptional silencing such as Mi-2/NuRD complex, which contains histone deacetylases (Bogdanovic and Veenstra, 2009). The role of DNA methylation is a combination of interfering with transcription at repressed genes, and modifying chromatin structure through its interplay with histone-modifying enzymes (Rose and Klose, 2014).

Removal of DNA methylation occurs through successive oxidation of the 5mC into 5-carboxylcytosine (5caC) through 5-hydroxymethylcytosine (5hmC) and 5-formylcytosine (5fC) by members of the TET family, followed by conversion to cytosine through base-excision repair (Figure 1.2.1) (Wu and Zhang, 2015, Zhang et al., 2012). While initially thought to be part of a pathway to remove 5mC, it is now emerging that these DNA modifications may be part of the epigenetic code as these modifications are not distributed randomly in the genome and at least 5hmC is a stable mark. However sequencing of these marks is technically challenging owing to their low abundance and the biochemistry involved (Mellen et al., 2012). Interestingly, DNA containing 5fC has properties unlike A- and B-DNA, and this has been called F-DNA. This form of DNA has a much narrower and deeper major groove, and much wider and shallower minor groove compared to B-DNA (Raiber et al., 2015). 5hmC has already been called the ‘sixth DNA base’, and it is quite possible that 5fC and 5caC reflect seventh and eighth DNA bases.

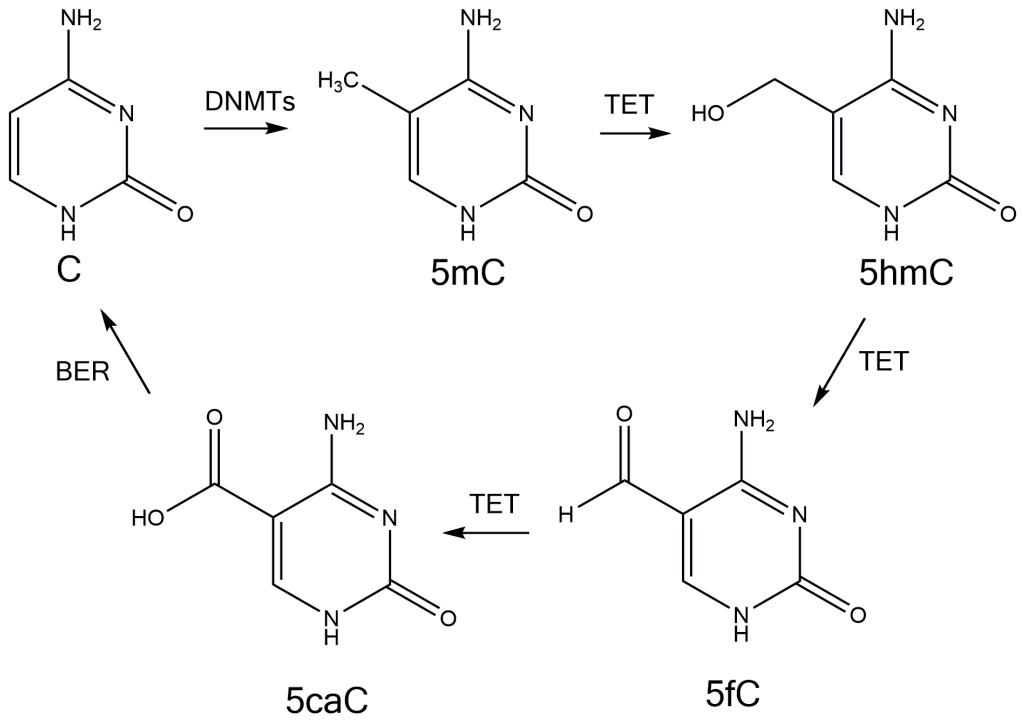

Figure 1.2.1. Conversion from cytosine to 5mC occurs through DNMTs. Conversion between 5mC, 5hmC, 5fC and 5caC is carried out by TET enzymes. Regeneration of cytosine occurs through Base Excision Repair (BER).

A second level of control for gene expression is the modification of the histone tails in the context of nucleosomes. These histones can contain a number of different modifications, such as methylation, acetylation, phosphorylation, and ubiquitination, among others (Patel and Wang, 2013). Acetylation of lysines in histone tails is carried out through the action of Histone Acetyl Transferases (HATs). This modification neutralises the positive charge on the lysine residues and weakens DNA-histone interactions (Hong et al., 1993). This makes chromatin more accessible and is associated with transcriptional activation, either by facilitating transcription factor binding through more accessible chromatin, or recruitment of transcriptional coactivators. Acetylated histones are recognised by a number of domains, notably bromodomains, which can occur in tandem to enhance their binding (Patel and Wang, 2013, Jacobson et al., 2000). Structurally, these conserved domains contain four alpha helices, which bind histones using a hydrophobic pocket formed by two loops (ZA and BC) and two helices (B and C) (Figure 1.2.2) (Poncet-Montange et al., 2015)

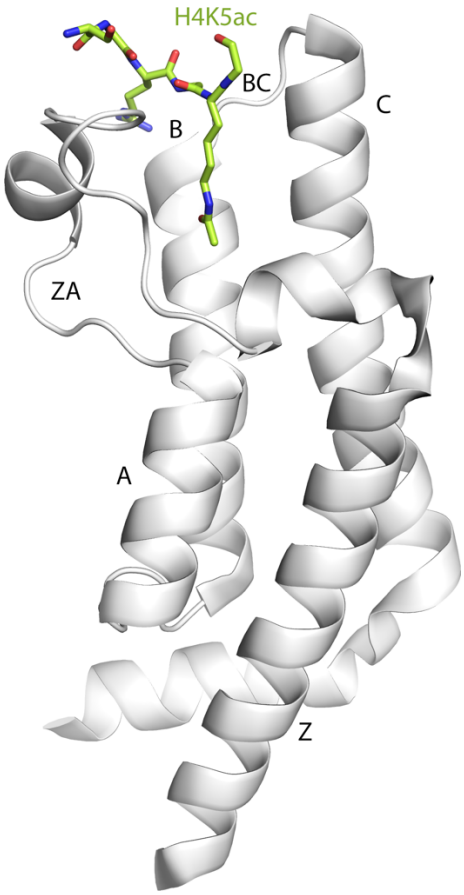

Figure 1.2.2. H4K5Ac bound to the bromodomain of ATAD2 (PDB: 4TT2).

Removal of histone acetylation is performed by Histone Deacetylases (HDACs), although these proteins are also able to deacetylate other types of proteins. There are three classes of HDACs. Class I and II require a Zn^{2+} ion for their catalytic activity, while class III HDACs are NAD+-dependent. Class I HDACs contain HDAC 1, 2, 3, and 8, and are expressed ubiquitously, while class II, containing HDAC 4, 5, 6, 7, 9a, 9b, 9c (catalytically inactive), and 10, are thought to be more tissue-specific. HDACs are identified as class I or class II based on their sequence identity to yeast HDACs RPD3 and HDA1, respectively. Another Zn^{2+}-dependent HDAC, HDAC 11, has little sequence homology to either RPD3 or HDA1. Class I HDACs are almost exclusively nuclear, although HDAC3 is found in both the cytosol and nucleus due to the presence of a nuclear localisation signal as well as a nuclear export signal (de Ruijter et al., 2003). There are other co-factors that are required for the activation of HDACs, for example the interaction of HDAC3 with the NCoR/SMRT Deacetylase Activation Domain is required for its activation, and it is dependent on inositol-tetraphosphate (IP4) (Watson et al., 2012). Inhibition of HDACs has been successful in cancer therapy, although the exact mechanism by which these HDAC inhibitors work is unclear. However the effects are a combination of direct

HDAC inhibition and disruption of HDAC containing complexes, which lead to gene expression changes through both impaired histone and transcription factor deacetylation (Delcuve et al., 2012). In neurons, deacetylation of histone tails by HDAC3 is dependent on MeCP2 recruiting NCoR/SMRT corepressor complex. This interaction is disturbed in RTT, and will be discussed in next sections.

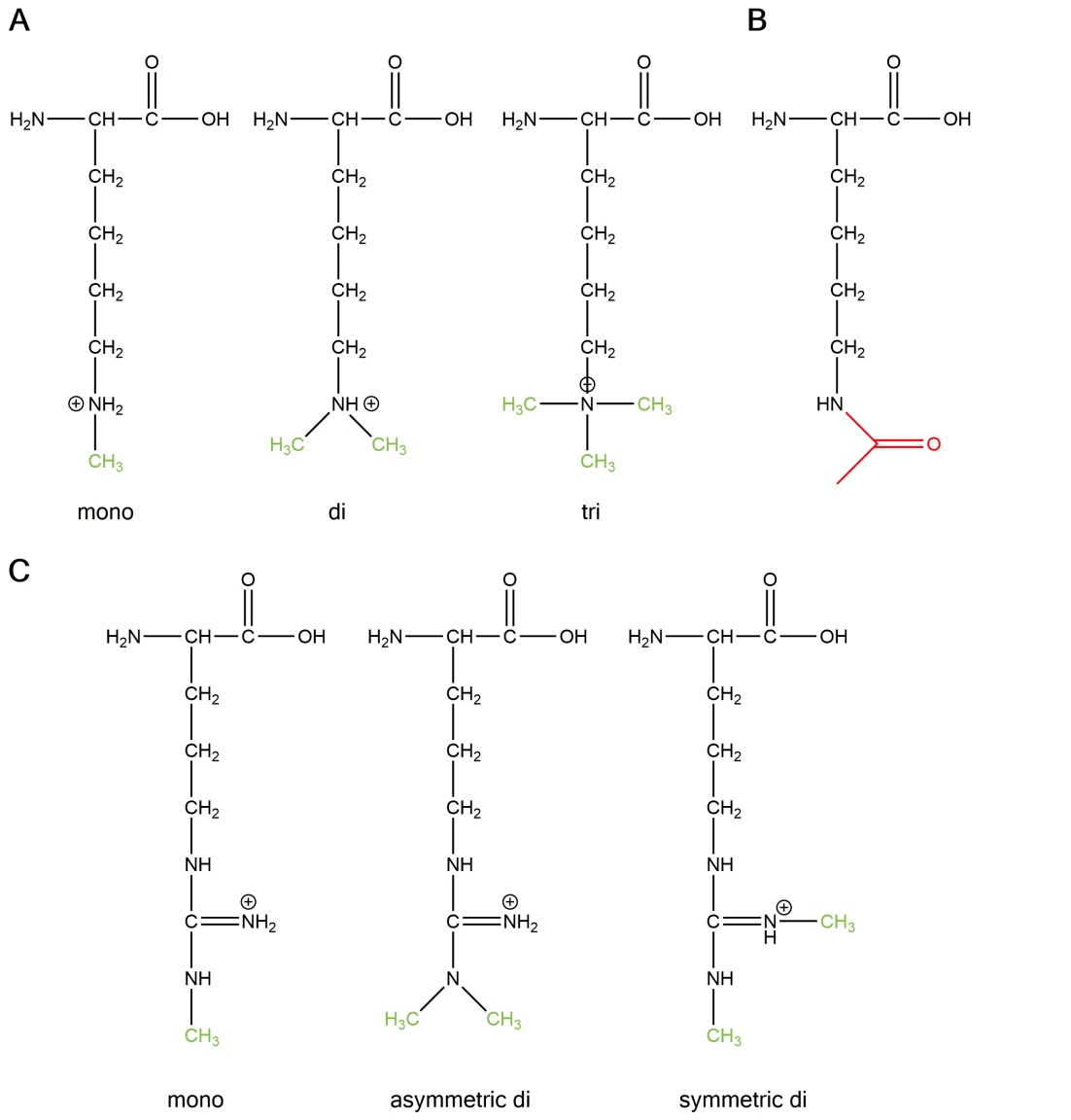

Figure 1.2.3. A) Possible methylation states of lysine. B) Structure of acetyllysine. Acetyl moiety has been colored red. C) Possible methylation states of arginine.

Methylation occurs on lysine and arginine residues, and is carried out by histone lysine methyltransferases or arginine methyltransferases (Bannister and Kouzarides, 2011). Demethylation of both arginine and lysine residues occurs through oxidation of the methyl group and subsequent release of formaldehyde (Walport et al., 2012). Methylation gives rise to a plethora of combinations, as lysines can be mono-, di-, and trimethylated, while arginine can mono- or dimethylated. In addition, arginine modifications can be either symmetrical or asymmetrical (Figure 1.2.3). The role of methylation is more complex, as for example monomethylation of H3K9 and H3K27 coincides with active genes, while trimethylation of H3K9 and H3K27 are associated with repressed genes (Barski et al., 2007). Methylation of lysines is recognised by, for example, chromodomains, PHD finger domains, and Tudor domains, the latter being also involved in binding to methylated arginines (Gayatri and Bedford, 2014). In addition, some WD40 domains are involved in binding to methylated histones, as described in section 1.4.1.

While the number of different modifiable histone residues is high, further combinations can occur through histones modified with more than one chemical mark, where different modifications can act antagonistically. An example of this is the binding of HP1 to H3K9me3, where phosphorylation of S10 on histone 3 (H3pS10) abolishes the binding of HP1 (Fischle et al., 2005). In this light, due to the sheer amount of possible combinations of histone modifications, arising from the number of modifications plus the number of modifiable residues on histone tails, the histone code has not yet been fully elucidated.

In this thesis I will discuss a member of the methyl-CpG-binding protein family, MeCP2, which mediates control over gene transcription through both DNA methylation and histone modifications.

1.3 Methyl-CpG-binding Protein 2 (MeCP2)

1.3.1 MeCP2 domain organization

Interest in MeCP2 function was sparked in 1999 when it was found that mutations in MeCP2 cause Rett syndrome (Amir et al., 1999). Since then, it has been proposed that the role of MeCP2 extends beyond that of just binding to methylated DNA, and that it is very complex, being potentially involved in multiple protein-protein as well as protein-DNA interactions. Because of this, several annotations of the MeCP2 protein exist. However, in this thesis the discussion will be based on the annotation shown in Figure 1.3.1 as this annotation is based on the two domains that contain a number of residues known to be mutated in RTT.

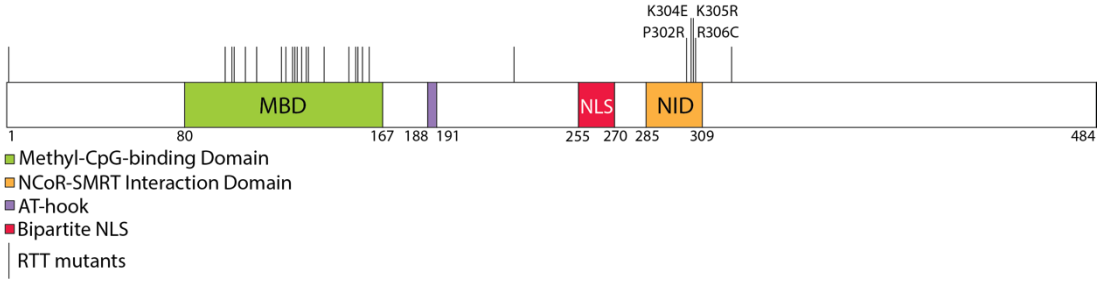

Figure 1.3.1. Mouse MeCP2 isoform A domain organization. The missense mutations in the NID that cause RTT have been highlighted. All of the mutations shown are present in RTT patients but not in their parents (Lyst et al., 2013)

Primary sequence analysis indicates that the majority of MeCP2 is natively disordered. Furthermore, circular dichroism analysis of two fragments of MeCP2, amino acids 78-162 and 205-310 showed that these are 40% and 85% unstructured, respectively (Adams et al., 2007). The first fragment corresponds to the methyl-CpG-binding Domain (MBD), while the second fragment was originally annotated as the transcriptional repression domain (TRD) due to its association with Sin3a and HDAC1/2 (Nan et al., 1998), however in light of evidence that MeCP2 is not part of the Sin3a complex and does not stably associate with it (Klose and Bird, 2004), this annotation is no longer widely used.

MeCP2 protein has two well-characterised domains called the methyl-CpG-binding Domain (MBD) and the NCoR/SMRT Interaction Domain (NID). These two domains mediate the interactions with DNA and NCoR/SMRT co-repressor complex components TBL1X/TBL1XR1, respectively (Nan et al., 1993, Lyst et al., 2013). Between the MBD and NID is a linker region that contains an AT-hook (called AT-hook 1), which is a protein motif that binds to AT-rich DNA, however its precise function in MeCP2 is unknown (Klose et al., 2005).

Additionally, there is a potential bipartite nuclear localisation signal (NLS). Unpublished data by Dr. Matthew Lyst in Adrian Bird's lab in Edinburgh has shown that the NLS is not required for the nuclear localisation of MeCP2, which can be achieved by a functioning MBD. However, MeCP2 containing a mutated MBD is able to localise to the nucleus if the NLS present, while MeCP2 with both mutated MBD and NLS stays in the cytoplasm. There are additional reports suggesting that the NLS is another AT-Hook (AT-hook 2) (Baker et al., 2013), although the sequence is not the canonical consensus sequence of HMG1 AT-hooks (Banks et al., 1999) (Figure 1.3.2). There are two nonsense mutations affecting this region, R270X and G273X, which cause RTT with varying severity and survival rates, and it has been proposed that the disruption of this AT-hook 2 is the basis for the more severe phenotype of R270X compared to G273X (Baker et al., 2013). However, no RTT missense mutations have been reported in these regions and it is unclear whether these contribute specifically to RTT phenotype, or whether they cause RTT due to deletion of a large part of the MeCP2 protein, including the NID.

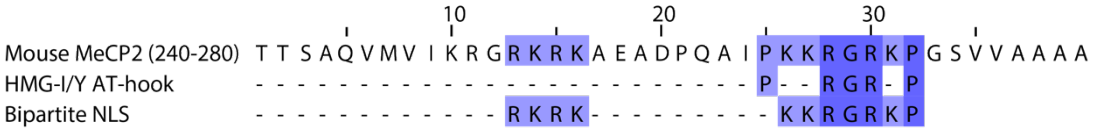

Figure 1.3.2. Sequences of mouse MeCP2 residues 240-280 with HMG-I/Y AT-hook and potential bipartite NLS.

Despite the fact that the vast majority of *de novo* missense RTT mutations cluster in these two domains, with only three of them appearing outside the domain boundaries, the majority of the protein is conserved in vertebrates (Figure 1.3.3). On one hand, this implies that these regions may have functional significance, either by mediating protein-protein or protein-DNA interactions, or by ensuring the correct folding and stability of MeCP2. On the other hand, there is large number of neutral variants present in these interdomain regions (Lyst and Bird, 2015). As MeCP2 is largely unstructured, and therefore the interdomain regions are likely not required for the stability of the three-dimensional fold, it is possible that the expansion of the protein is neutral and that these regions have no true functional significance. However, the C-terminal region has an interesting characteristic, namely a conserved polyhistidine run between amino acids 366-372 (numbering in humans).

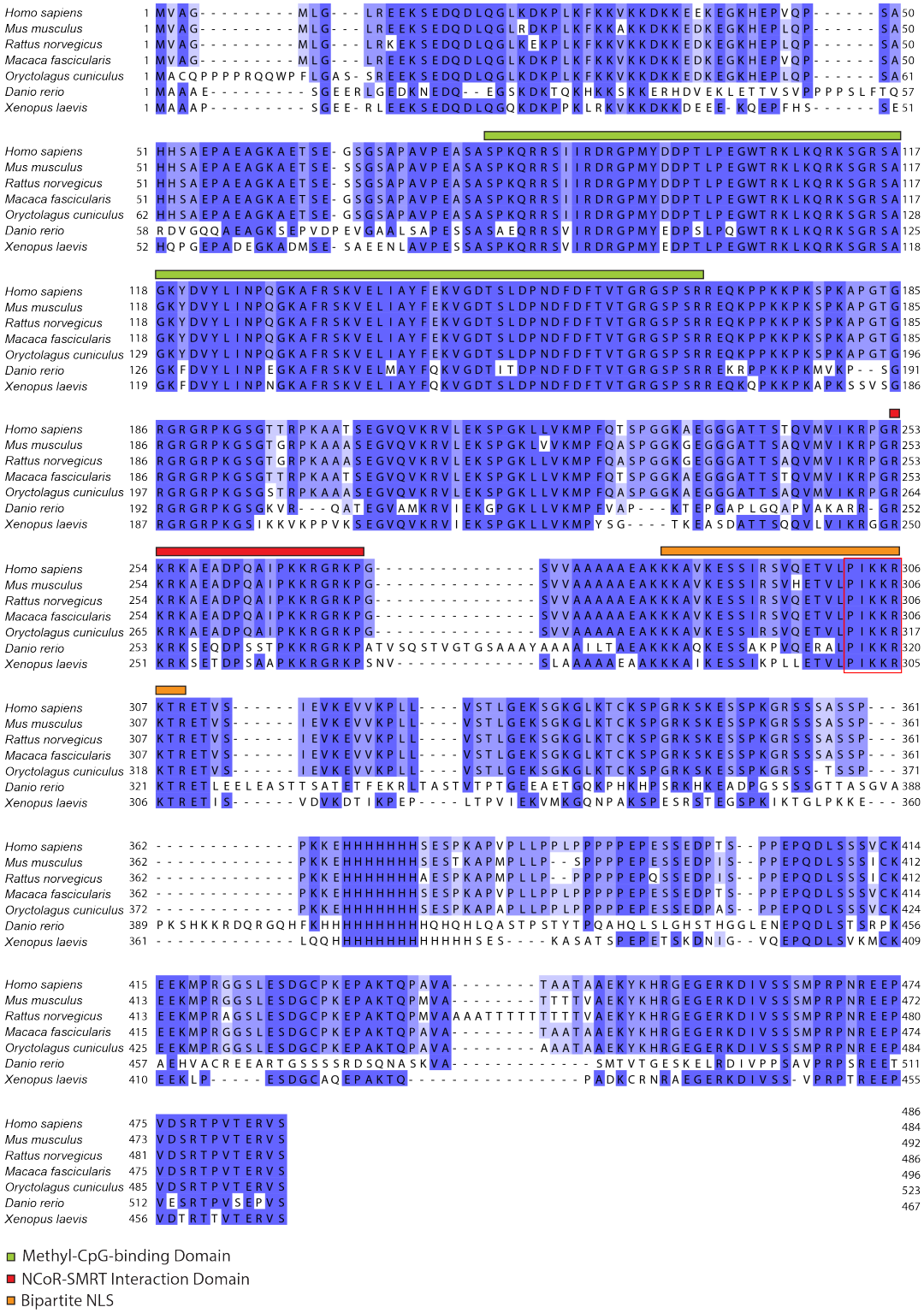

Figure 1.3.3. Sequence alignment of MeCP2 from various vertebrates. Sequences are the canonical sequences from Uniprot (isoforms beta). The region in the NID containing RTT mutations has been outlined with a red box. Species are as follows: *Homo sapiens* – human, *Mus musculus* – house mouse, *Rattus norvegicus* – brown rat, *Macaca fascicularis* – crab-eating macaque, *Oryctolagus cuniculus* – European rabbit, *Danio rerio* – zebrafish, *Xenopus laevis* – African clawed frog.

A study that analysed human proteins containing polyhistidine runs found that many of the proteins identified were associated with brain or neuronal function. Furthermore, polyhistidine runs were proposed to be a novel signal for targeting proteins to nuclear speckles, which are regions involved in RNA splicing (Salichs et al., 2009). As the C-terminal domain of MeCP2 has been implicated in alternative splicing by associating with YB-1 (Young et al., 2005), it is possible that the polyhistidine run is involved in this function.

1.3.2 MBD and its role in RTT

As previously mentioned, MeCP2 was identified based on its ability to bind to methylated DNA. A crystal structure of the MeCP2 MBD bound to methylated DNA shows that the residues in the MBD (R111 and R133) do not recognise the methylated cytosines (5mC) directly, but rather through ordered water molecules in the major groove (Figure 1.3.4) (Ho et al., 2008). The RTT mutations in the MBD either disrupt the binding to (hydroxy)methylated DNA, or affect the stability of the protein (Ghosh et al., 2008, Brown et al., 2016, Mellen et al., 2012).

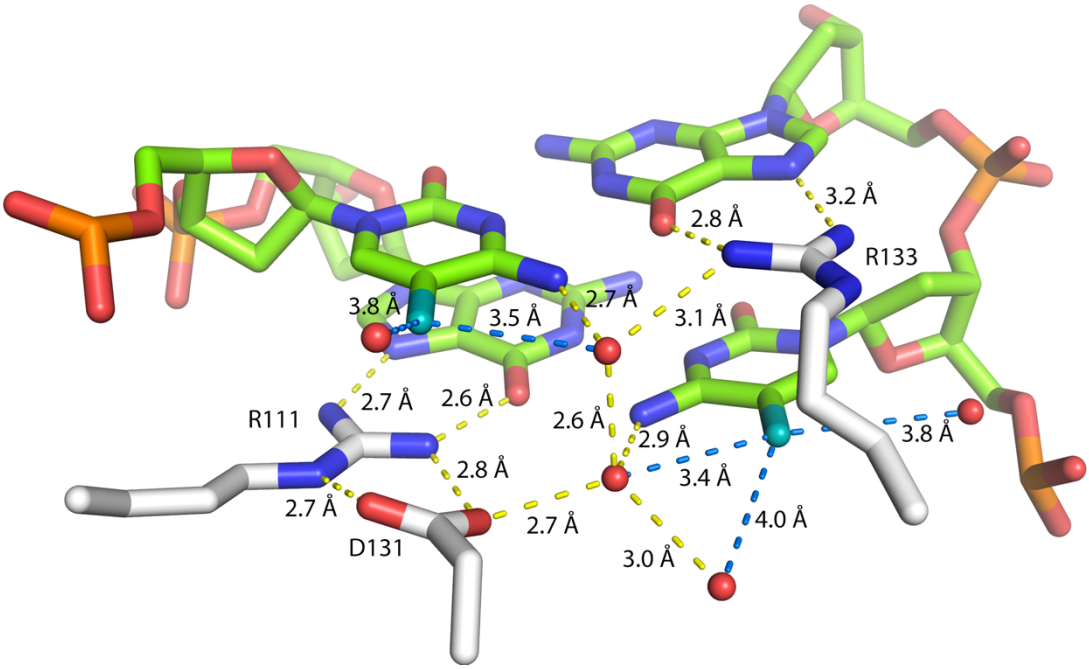

Figure 1.3.4. Recognition of methylated cytosines by the MBD of MeCP2. Ordered waters, shown as red spheres, are present in the major groove of the DNA. Methyl groups have been coloured teal. CH…O hydrogen bonds have been shown as blue lines.

MeCP2 has been shown to be highly abundant in neurons, with genome-wide ChIP data showing approximately one molecule of MeCP2 per two mCpGs. These levels reach those of histone octamers. In addition, it binds throughout the genome and tracks methylated CpG

density (Skene et al., 2010). This is supported by an observation that while the affinity of MeCP2 for unmethylated DNA is only three times lower than monomethylated DNA, the presence of multiple CpG repeats increases its affinity for methylated DNA (Fraga et al., 2003). It was therefore proposed that MeCP2 acts as a global repressor of transcription (Skene et al., 2010). However, later it was shown that MeCP2 acts not only as a repressor, but also as an activator through its interaction with CREB1 (Chahrour et al., 2008), indicating that its role in gene transcription is more complex.

In addition to 5mC binding, MeCP2 also binds 5-hydroxymethylcytosine (5hmC), another epigenetic mark which is enriched in brain. The R133C mutation abolishes the binding of MeCP2 to 5hmC, while retaining its ability to bind 5mC (Mellen et al., 2012). MeCP2 overexpression decreased the levels of 5hmC, while loss of MeCP2 showed increased levels of 5hmC. Consistent with this observation, MeCP2 was shown to be able to block TET enzymes from converting 5mC into 5hmC *in vitro* (Szulwach et al., 2011). Furthermore, MeCP2 has been shown to bind preferentially to long genes with high mCA content and repress them, whereas short genes with low or high mCA content, and long genes with low mCA content are not misregulated in an MeCP2 knockout. It was demonstrated that long genes are more likely to be associated with neuronal function (Gabel et al., 2015). Additionally, neurons have been shown to have much higher level of mCH methylation (where H is any of the A, T or C bases) compared to other cell types (Kinde et al., 2015). In view of all this, it has been proposed that RTT can arise from brain-specific misregulation of long genes (Gabel et al., 2015).

1.3.3 NID and its role in RTT

MeCP2 represses transcription through its interaction with the NCoR/SMRT co-repressor complex (Figure 1.3.5) (Lyst et al., 2013). The NCoR/SMRT complex is made up by a scaffold consisting of of TBL1X or its paralogue TBL1XR1, onto which NCoR/SMRT and GPS2 assemble. HDAC3 interacts with a Deacetylase Activating Domain (DAD) on NCoR/SMRT, completing the complex (Oberoi et al., 2011) (Figure 1.3.5).

Several missense mutations in the NID have been shown to cause RTT (Figure 1.3.1). In particular, it has been shown that MeCP2 R306C mutation disrupts the interaction with TBL1X, and by extension with NCoR/SMRT, both *in vivo* and *in vitro* (Lyst et al., 2013). It has also been shown that MeCP2 is phosphorylated at T308 in response to neuronal activity and that this activity disrupts the interaction of MeCP2 with NCoR/SMRT. Mutations that mimic the phosphorylation state (T308D/T308E) show decreased levels of transcriptional

repression compared to the wild-type, consistent with decreased levels of association of MeCP2 with TBL1X. Mice carrying a T308A mutation showed deficiencies in the regulation of neuronal activity-dependent genes. The T308A mouse model also had a RTT-like phenotype although it was not as severe as R306C. Therefore both the loss of function (R306C) and non-phosphorylatable form of MeCP2 (T308A) can lead to RTT-like symptoms in mice (Ebert et al., 2013).

The region spanning the NID has been also annotated as a 'basic patch', owing to the high abundance of positively charged amino acids. There have been reports that this region, including the part of the protein including the NID, is involved in DNA binding (Heckman et al., 2014), and that potentially it is the loss of this interaction that leads to RTT. This leaves open the possibility that the NID/'basic patch' has multiple functions, and that mutations in the NID cause RTT by abolishing the interaction of MeCP2 with DNA and/or NCoR/SMRT. This will be addressed in chapter five.

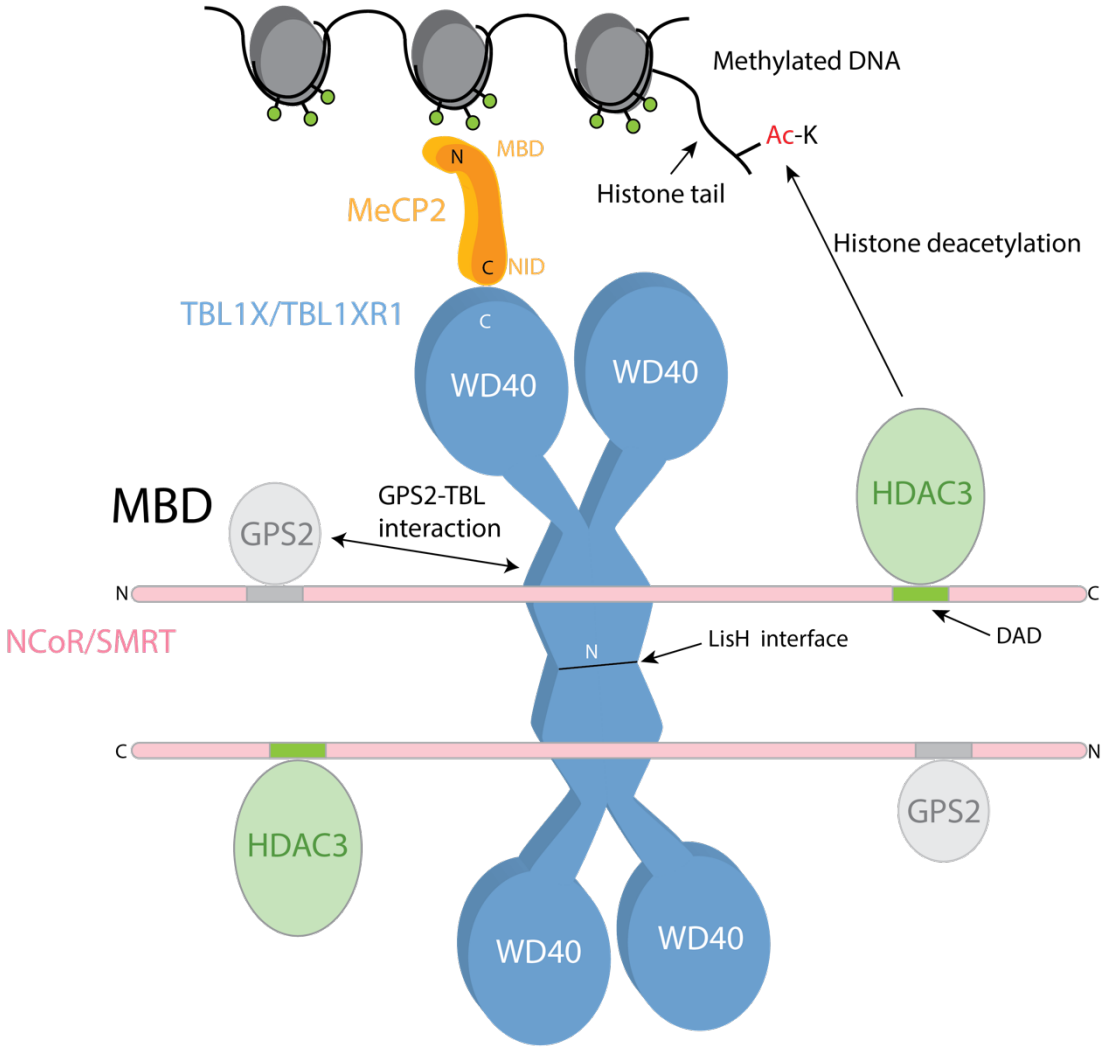

Figure 1.3.5. Schematic overview of NCoR/SMRT corepressor complex. TBL1X and/or TBL1XR1 (blue) form a tetrameric scaffold through their LisH domains. This scaffold interacts with the NCoR/SMRT (pink) and GPS2 (grey). In addition, MeCP2 NID (orange) interacts with WD40 domain of TBL1X/TBL1XR1. Currently it is unknown what the stoichiometry of this interaction is. This complex is recruited to DNA via the MeCP2 MBD, which binds to methylated DNA (green circles). Binding of HDAC3 to the Deacetylase Activation Domain (DAD) of NCoR/SMRT activates it and leads to deacetylation of histone tails. N and C stand for N- and C-terminus, respectively.

The fact that MeCP2 interacts with the NCoR/SMRT co-repressor complex has given insight into how MeCP2 might act as a transcriptional regulator, bridging methylated DNA and NCoR/SMRT, and leading to deacetylation of histone tails by recruitment of HDAC3 (Figure 1.3.5 and Figure 1.3.6). This model is consistent with an experiment that showed levels of H3 acetylation rise in mice with $MeCP2^{R309X}$ mutant (Shahbazian et al., 2002, Skene et al., 2010). While this mutation does not cause a severe RTT phenotype, MeCP2 containing this mutation shows reduced association with HDAC3 (Lyst et al., 2013), possibly indicating a weakened interaction between MeCP2 and NCoR/SMRT complex.

Recently, the effect of HDAC3 knockout in neurons was investigated. Mice with HDAC3 conditional knockout (KO) in forebrain neurons showed RTT-like symptoms and misregulation (both up and downregulation) of over 300 genes, many of them associated with neuronal function. There were overlaps of misregulated genes between HDAC3 and MeCP2 knockout neurons, indicating that the transcription of these genes might be linked. Indeed, MeCP2 targeted HDAC3 to the promoters of active genes as shown by a combination of ChIP-seq and qPCR. Genes found to be downregulated in HDAC3 KOs were associated with a transcription factor called FOXO (Nott et al., 2016). It has been previously demonstrated that acetylation of FOXO proteins by CBP/p300 leads to their diminished binding to DNA, and subsequent dissociation (Daitoku et al., 2011). In HDAC3 KOs, levels of FOXO acetylation rose, providing a rationale as to why HDAC3 loss can cause downregulation of certain genes (Nott et al., 2016). Interestingly, MeCP2 dissociation from the BDNF promoter is SIRT1-dependent (an HDAC), and depends on the deacetylation of K464 of MeCP2 (in mouse isoform alpha, K449 in human isoform beta) (Zocchi and Sassone-Corsi, 2012).

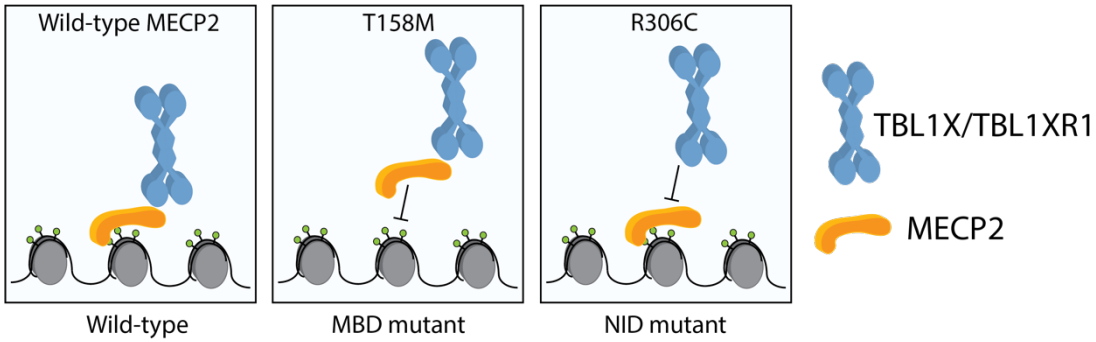

Figure 1.3.6. Proposed model of the role of MeCP2 mutants. In neurons containing the wild-type MeCP2, MeCP2 interacts with methylated DNA (green circles) and with TBL1X/TBL1XR1 (blue). In MBD mutations (example - T158M), MeCP2 is unable to interact with DNA while still maintaining its binding to TBL1X/TBL1XR1. In NID mutations (example - R306C), MeCP2 can bind to DNA but is unable to recruit TBL1X/TBL1XR1.

Given the roles of both acetylation and HDACs in regulating gene transcription, it is unsurprising that abolition of the MeCP2-NCoR/SMRT complex interaction by NID RTT mutations causes misregulation of gene transcription (Ballestar et al., 2005). However, the exact role of HDAC3 in RTT pathology has not been elucidated. Recent evidence suggests that RTT may be caused by disrupted deacetylation of transcription factors and MeCP2. However, given the role of MeCP2 at the centre of transcription control, and the relative non-specificity of HDACs, the role of HDAC3 is likely to be as complex as that of MeCP2. It is possible that the RTT pathology in the case of NID mutations arises not only from the failure to recruit HDAC3, but also from the loss of binding to TBL1X/TBL1XR1, as will be discussed in the experimental work present in this thesis.

1.4 Transducin-beta-like 1X (TBL1X) and Transducin-beta-like 1X-related (TBL1XR1)

TBL1X and TBL1XR1, the two proteins that make up the scaffold of NCoR/SMRT complex, are shown in Figure 1.4.1. They are paralogous proteins that have two domains; an N-terminal LisH (LIS1-homology) domain, and a C-terminal eight-bladed WD40 domain, which comprises most of the protein. The roles of WD40 domains are discussed in section 1.4.1.

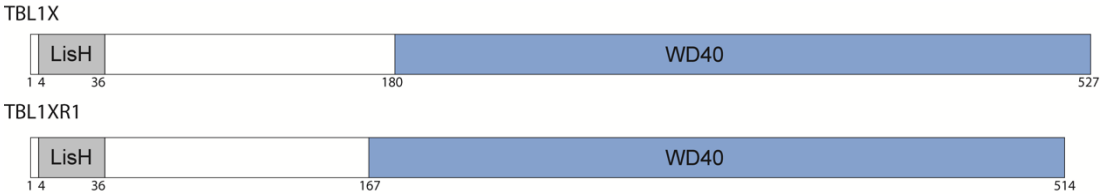

Figure 1.4.1. Domain organization of isoform 2 of mouse TBL1X and TBL1XR1. Domain annotations are based on Uniprot.

The LisH motif is generally involved in mediating protein-protein interactions and is made up of two alpha helices (Emes and Ponting, 2001). It is involved in both dimerisation and tetramerisation of TBL1X and TBL1XR1. Since the LisH motif is conserved between TBL1X and TBL1XR1, it is possible that the proteins can form dimers of homodimers, and possibly heterodimers, however this has not been demonstrated experimentally. The region of human TBL1X spanning residues 1-70 (identical to mouse TBL1X) interacts with GPS2 (G-protein pathway Suppressor 2) and NCoR/SMRT (Oberoi et al., 2011). Interaction of GPS2 with TBL1X is required to protect it from degradation by the proteasome, and this is mediated by the methylation of GPS2 (Huang et al., 2015). Mutation in TBL1XR1 (G70D) has been shown to cause West syndrome, which causes mental retardation and RTT-like phenotype (Saitsu et al., 2014). It is unknown whether this mutation affects the interaction of TBL1XR1 with GPS2, with NCoR, or whether it has a distinct effect.

Both TBL1X and TBL1XR1 also interact with histone H4, and preferentially with hypoacetylated H4 (Yoon et al., 2005), although this interaction has not been mapped. Whether the binding of TBL1X to hypoacetylated H4 and the recruitment of HDAC3 are separate, or whether deacetylation of histone tails by HDAC3 reinforces NCoR/SMRT complex binding through TBL1X/TBL1XR1 has not been addressed.

According to The Human Protein Atlas, both TBL1X and TBL1XR1 are expressed in every tissue tested and found predominantly in nuclei (The Human Protein Atlas). They have, however, distinct roles in cells by interacting with a different subset of nuclear receptors, although some overlap exists (Perissi et al., 2004). A third protein called TBL1Y is present on the Y chromosome which is unable to repress promoter activity that TBL1X and TBL1XR1 repress, suggesting a role separate from both TBL1X and TBL1XR1 (Yan et al., 2005). As these three proteins are conserved, save for a short polyalanine region (Figure 1.4.2), it is not immediately clear what underlies these different behaviours. Interestingly, the appearance of polyalanine stretches has been proposed to occur by convergent evolution in proteins that

share functions (Lavoie et al., 2003). It is therefore possible that the non-conserved polyalanine stretch confers specificity to these three proteins.

```
TBL1X isoform 1     1  MTELAGASSSCCHRPAGRGAMQSVLHHFQRLRGREGGSHFINTSSPRGEA  50
TBL1X isoform 2        --------------------------------------------------
TBL1Y                  --------------------------------------------------
TBL1XR1                --------------------------------------------------

TBL1X isoform 1    51  KMSITSDEVNFLVYRYLQESGFSHSAFTFGIESHISQSNINGTLVPPAAL  100
TBL1X isoform 2     1  -MSITSDEVNFLVYRYLQESGFSHSAFTFGIESHISQSNINGTLVPPAAL  49
TBL1Y               1  -MSITSDEVNFLVYRYLQESGFSHSAFTFGIESHISQSNINGTLVPPSAL  49
TBL1XR1             1  -MSISSDEVNFLVYRYLQESGFSHSAFTFGIESHISQSNINGALVPPAAL  49

TBL1X isoform 1   101  ISILQKGLQYVEAEISINEDGTVFDGRPIESLSLIDAVMPDVVQTRQQAF  150
TBL1X isoform 2    50  ISILQKGLQYVEAEISINEDGTVFDGRPIESLSLIDAVMPDVVQTRQQAF  99
TBL1Y              50  ISILQKGLQYVEAEISINKDGTVFDSRPIESLSLIVAVIPDVVQMRQQAF  99
TBL1XR1            50  ISIIQKGLQYVEAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQQAY  99

TBL1X isoform 1   151  REKLAQQQASAAAAAAAATAAATAATTTSAGVSHQNPSKNREATVNGEEN  200
TBL1X isoform 2   100  REKLAQQQASAAAAAAAATAAATAATTTSAGVSHQNPSKNREATVNGEEN  149
TBL1Y             100  GEKLTQQQASAAATEAS--AMAKAATMTPAAISQQNPPKNREATVNGEEN  147
TBL1XR1           100  RDKLAQQQAAAAAAAAAA-------------ASQQGSAKNGENTANGEEN  136

TBL1X isoform 1   201  RAHSV-NNHAKPMEIDGEVEIPSSKATVLRGHESEVFICAWNPVSDLLAS  249
TBL1X isoform 2   150  RAHSV-NNHAKPMEIDGEVEIPSSKATVLRGHESEVFICAWNPVSDLLAS  198
TBL1Y             148  GAHEI-NNHSKPMEIDGDVEIPPNKATVLRGHESEVFICAWNPVSDLLAS  196
TBL1XR1           137  GAHTIANNHTDMMEVDGDVEIPPNKAVVLRGHESEVFICAWNPVSDLLAS  186

TBL1X isoform 1   250  GSGDSTARIWNLNENSNGGSTQLVLRHCIREGGHDVPSNKDVTSLDWNTN  299
TBL1X isoform 2   199  GSGDSTARIWNLNENSNGGSTQLVLRHCIREGGHDVPSNKDVTSLDWNTN  248
TBL1Y             197  GSGDSTARIWNLNENSNGGSTQLVLRHCIREGGHDVPSNKDVTSLDWNSD  246
TBL1XR1           187  GSGDSTARIWNLSENSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSE  236

TBL1X isoform 1   300  GTLLATGSYDGFARIWTEDGNLASTLGQHKGPIFALKWNRKGNYILSAGV  349
TBL1X isoform 2   249  GTLLATGSYDGFARIWTEDGNLASTLGQHKGPIFALKWNRKGNYILSAGV  298
TBL1Y             247  GTLLAMGSYDGFARIWTENGNLASTLGQHKGPIFALKWNKKGNYVLSAGV  296
TBL1XR1           237  GTLLATGSYDGFARIWTKDGNLASTLGQHKGPIFALKWNKKGNFILSAGV  286

TBL1X isoform 1   350  DKTTIIWDAHTGEAKQQFPFHSAPALDVDWQNNTTFASCSTDMCIHVCRL  399
TBL1X isoform 2   299  DKTTIIWDAHTGEAKQQFPFHSAPALDVDWQNNTTFASCSTDMCIHVCRL  348
TBL1Y             297  DKTTIIWDAHTGEAKQQFPFHSAPALDVDWQNNMTFASCSTDMCIHVCRL  346
TBL1XR1           287  DKTTIIWDAHTGEAKQQFPFHSAPALDVDWQSNNTFASCSTDMCIHVCKL  336

TBL1X isoform 1   400  GCDRPVKTFQGHTNEVNAIKWDPSGMLLASCSDDMTLKIWSMKQEVCIHD  449
TBL1X isoform 2   349  GCDRPVKTFQGHTNEVNAIKWDPSGMLLASCSDDMTLKIWSMKQEVCIHD  398
TBL1Y             347  GCDHPVKTFQGHTNEVNAIKWDPSGMLLASCSDDMTLKIWSMKQDACVHD  396
TBL1XR1           337  GQDRPIKTFQGHTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHD  386

TBL1X isoform 1   450  LQAHNKEIYTIKWSPTGPATSNPNSNIMLASASFDSTVRLWDIERGVCTH  499
TBL1X isoform 2   399  LQAHNKEIYTIKWSPTGPATSNPNSNIMLASASFDSTVRLWDIERGVCTH  448
TBL1Y             397  LQAHSKEIYTIKWSPTGPATSNPNSSIMLASASFDSTVRLWDVEQGVCTH  446
TBL1XR1           387  LQAHNKEIYTIKWSPTGPGTNNPNANLMLASASFDSTVRLWDVDRGICIH  436

TBL1X isoform 1   500  TLTKHQEPVYSVAFSPDGKYLASGSFDKCVHIWNTQSGNLVHSYRGTGGI  549
TBL1X isoform 2   449  TLTKHQEPVYSVAFSPDGKYLASGSFDKCVHIWNTQSGNLVHSYRGTGGI  498
TBL1Y             447  TLMKHQEPVYSVAFSPDGKYLASGSFDKYVHIWNTQSGSLVHSYQGTGGI  496
TBL1XR1           437  TLTKHQEPVYSVAFSPDGRYLASGSFDKCVHIWNTQTGALVHSYRGTGGI  486

TBL1X isoform 1   550  FEVCWNARGDKVGASASDGSVCVLDLRK                        577
TBL1X isoform 2   499  FEVCWNARGDKVGASASDGSVCVLDLRK                        526
TBL1Y             497  FEVCWNARGDKVGASASDGSVCVLDL--                        522
TBL1XR1           487  FEVCWNAAGDKVGASASDGSVCVLDLRK                        514
```

Figure 1.4.2. Sequence conservation of human TBL proteins. TBL1X has two isoforms, shown here.

As mentioned earlier, MeCP2 has been shown to interact with NCoR/SMRT and the RTT R306C mutation in MeCP2 abolishes the nuclear localisation of TBL1X in *ex vivo* assays. Recently, unpublished experiments carried out by Dr. Matthew Lyst from the Bird lab in

Edinburgh showed that MeCP2 interacts with the WD40 domain of TBL1X, and that mutating a single amino acid (E364G) abolishes this interaction. This interaction, as well as the potential role of TBL1X and TBL1XR1 in RTT syndrome and related intellectual disability disorders will be discussed at length in chapter five. In addition to RTT, both TBL1X and TBL1XR1 are involved in a variety of cancers (Daniels et al., 2014, Ramadoss et al., 2011, Garg and Aggarwal, 2002). Interestingly, levels of TBL1X and TBL1XR1 have been found to be both up- and down-regulated in different cancers (Daniels et al., 2014, Wang et al., 2014). In support of this role, TBL1X has been shown to be required for the recruitment of NF-κB, a transcription factor that regulates the expression of many genes involved in tumorigenesis (Ramadoss et al., 2011), and whose abnormal activation is commonly found in cancers (Perkins, 2004).

Both TBL1X and TBL1XR1 are involved in ligand-dependent nuclear cofactor exchange by recruiting the proteasome to degrade NCoR, which leads to binding of coactivators and subsequent gene activation (Perissi et al., 2004). As they are at the heart of gene transcription, it is not surprising that mutations in or aberrant expression of these proteins can lead to a variety of diseases caused by misregulation of gene expression. In the light of this, the fact that the loss of recruitment of NCoR/SMRT corepressor complex to DNA causes misregulation of gene expression in RTT fits with the observed loss of binding of MeCP2 to TBL1X and TBL1XR1.

1.4.1 WD40 domains

WD40 domains in eukaryotic organisms are extremely common, being among the ten most common domains in *Homo sapiens*, *Drosophila melanogaster*, and *Saccharomyces cerevisiae*. They rarely occur on their own, rather they are frequently present in conjunction with domains such as F-Box, LisH, and RING (Stirnimann et al., 2010), as is the case with TBL1X/TBL1XR1. The WD40 domains are made up of variable number of repeats or blades, depending on classification. A blade is formed from four beta strands. In a WD40 repeat, the repeating unit consists of beta strands d'-a-b-c (Figure 1.4.3), where strands a, b and c form part of one blade, while strand d' complements the beta sheet of a preceding blade. Consequently, a WD40 blade consists of beta strands a-b-c-d that bridge two repeats (Figure 1.4.3). To avoid confusion, only the term 'blade' will be used in this thesis. These approximately 40 amino acid long blades contain four antiparallel beta strands, which often end in a tryptophan-aspartate motif at the end of blade c, hence the name "WD40".

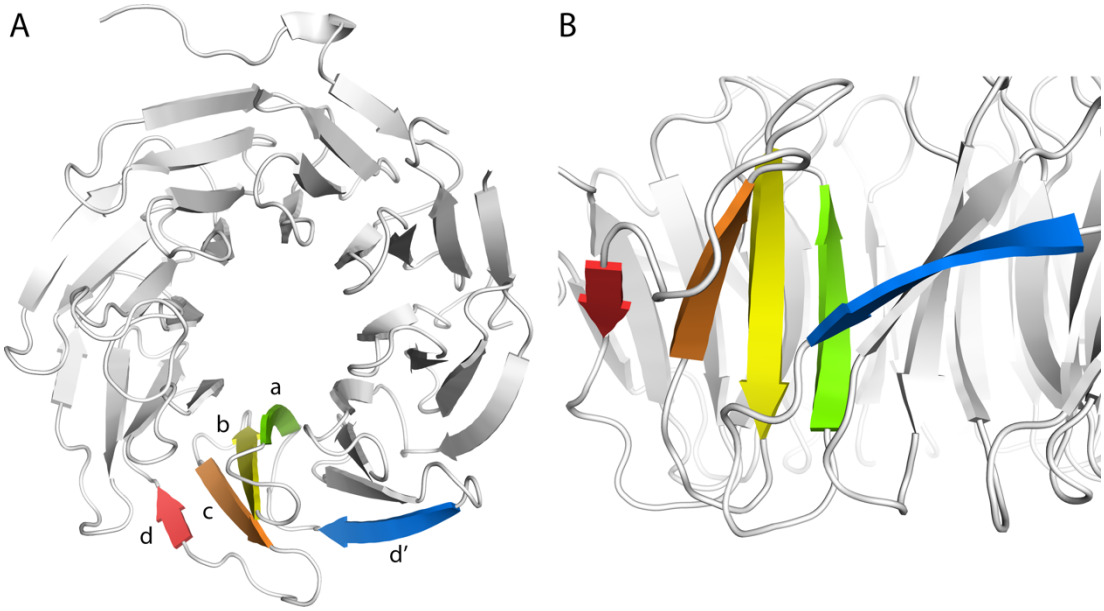

Figure 1.4.3. Cartoon representation of human TBL1XR1 WD40 domains (PDB: 4LG9). A) Beta strands which belong to one WD40 repeat have been labelled d', a, b, and c. Beta strands which belong to one WD40 blade are labeled a, b, c, and d. B) Close-up view of the same molecule.

WD40 domains are the largest family of beta-propeller proteins (Chen et al., 2011). Other families include the kelch family, YWTD family, NHL (containing YVTD motif), and YVTN family (Chaudhuri et al., 2008). The number of blades varies between these families; as of writing PDBe has records for four, five, six, seven, and eight-bladed beta propellers, as well as a ten-bladed sortilin, which has been annotated as WD40/YVTN-repeat-like protein. Analysis of WD40-repeat containing proteins in PDBe showed that usually the first beta strand at the N-terminus is the 'd' strand of the C-terminal blade, which completes the ring and provides structural stability, although that is not always the case as can be seen, for example, in DNA-damage binding protein 1 (DDB1) (PDB: 3E0C) (Figure 1.4.4). Furthermore, some WD40 domains are complemented by a blade from a different protein to finish the fold, as is the case with six-bladed Sec13, which uses a blade from Nup135C to complete a seven bladed WD40 fold (Whittle and Schwartz, 2010).

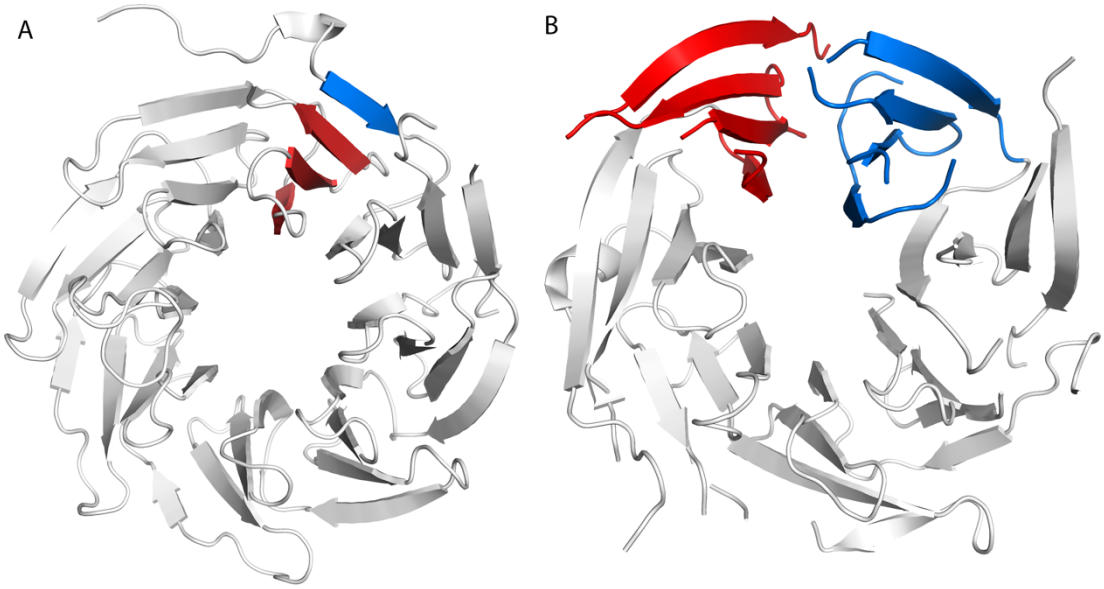

Figure 1.4.4. A) The first beta strand at the N-terminus on the protein (blue) forms a WD40 blade in conjunction with the three blades at the C-terminus of the domain (red) in human TBL1XR1. B) In DDB1, the domain closes through hydrophobic interactions between two blades instead.

While some beta-propellers are enzymes (Vandermarliere et al., 2009), no WD40 domain has been found to be catalytically active (Stirnimann et al., 2010). As such, WD40 domains are thought to be exclusively scaffolding proteins, serving as mediators of protein-protein and protein-DNA interactions. These interactions occur mostly through the top face, although bottom and side faces have been shown to mediate some interactions (Stirnimann et al., 2010). The variability in the interactions comes from these loops present between the d-a and b-c beta strands on the top face. Analysis of WD40 domains has found that there are certain amino acid positions in these blades and loops, called hotspots, that are often involved in mediating protein-protein interactions (Wu et al., 2012) (Figure 1.4.5). These amino acids in each blade are found at the opening of the central channel of the WD40 domain, providing an interaction surface. These residues will be discussed in detail in Chapter 5 in the context of MeCP2-TBL1XR1 interaction.

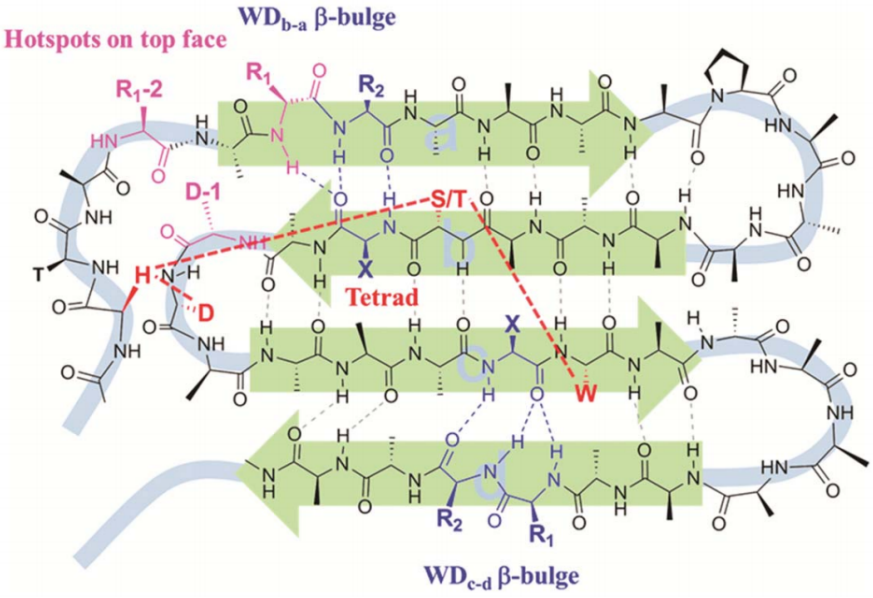

Figure 1.4.5. Structural representation of a single WD40 blade. Residues that are often found mediating protein-protein interactions are labelled R_1, R_1-2, and D-1 (pink). DH(S/T)W tetrad which has been highlighted here will be discussed in Chapter 5. (Figure taken from (Wu et al., 2012))

As mentioned previously, some WD40 domains are also involved in reading histone modification marks. For example, WDR5 recognises both H3K4me2 (Ruthenburg et al., 2006) and H3R2me2 (Migliori et al., 2012) Likewise, EED, which is part of the Polycomb Repressor Complex 2, recognises H3K27me3 (Margueron et al., 2009). In addition, these WD40 domains interact with lysine methyltransferases (KMTs), such as MLL1 in the case of WDR5 (Song and Kingston, 2008), and EZH2 in the case of EED (Han et al., 2007). In this way, these WD40 domains are able to link histone modifications with histone-modifying enzymes. In this light it is interesting that a variety of point mutations in TBL1XR1 WD40 domain cause intellectual disability, most likely due to its involvement in regulating gene expression in the brain through recruitment of HDAC3, although this mechanism does not involve a direct binding of the WD40 domain to histones. This will be discussed further in the next chapters.

It has been proposed that WD40 domains are involved in so many interactions as a protein family due to the ease with which they are able to evolve new binding sites. There is thought to be no defined folding order for the WD40 blades as mutations in individual blades do not affect the overall folding success, although they have an impact on the kinetics (Garcia-Higuera et al., 1998). Thus, point mutations in individual blades that can create new interaction surfaces are unlikely to disrupt the folding of the overall structure (Stirnimann et al., 2010). Because of their diverse nature and role as scaffolding proteins that are involved

in more than just binary interactions, WD40 domains are likely to be involved in a variety of diseases, and functional understanding of these proteins will uncover these roles.

1.5 Aims

The main aim of this project was to understand the molecular basis for the interaction between MeCP2 NID and TBL1X/TBL1XR1, and consequently understand why RTT mutations cause the loss of binding to TBL1X/TBL1XR1. This was achieved by co-crystallising and solving the structure of MeCP2 NID and TBL1XR1 using X-ray crystallography. In addition, the interaction was characterised using biophysical and biochemical methods.

In order to successfully complete the main aim, TBL1X/TBL1XR1 at high purity was required. To this end a protocol for large-scale expression and purification was developed. After successful purification, these proteins were characterised using biochemical and biophysical assays, both on their own and in conjunction with MeCP2 NID peptides. These experiments have provided information about the physical parameters affecting the interaction of these two proteins. Finally, MeCP2 NID peptide with TBL1X/TBL1XR1 was co-crystallised. The co-crystal structure gives a detailed understanding of how the RTT mutations can lead to the abolition of this interaction.

In addition to the main aim, a second aim was to examine the role of MeCP2 in linking DNA to NCoR/SMRT by probing whether MeCP2 binding to either DNA or NCoR/SMRT enhanced the affinity of the other component. For this, purification protocols for MeCP2 constructs of interest were developed. Subsequently, the affinities of MeCP2 towards DNA in the presence and absence of TBL1XR1 were investigated using several biophysical methods.

The following chapters in this thesis present the results of these experiments and answer most of these questions. Together they further the understanding of how MeCP2 NID mutations cause RTT.

2 Materials and Methods

2.1 Pre-existing constructs

Several constructs were cloned by Dr. Atlanta Cook prior to the start of the project for heterologous expression in *E. coli*, listed in Table 2.1.1.

Table 2.1.1. List of constructs cloned prior to the start of the project

Protein	Construct	Tag	Organism	Vector
MeCP2	197-339	N-GST	*H. sapiens*	pEC-KGC
MeCP2	266-339	N-GST	*H. sapiens*	pEC-KGC
MeCP2	286-339	N-GST	*H. sapiens*	pEC-KGC
TBL1X	141-527	N-GST	*H. sapiens*	pEC-KGC
TBL1X	159-527	N-GST	*H. sapiens*	pEC-KGC

2.2 Traditional cloning

Below are listed the constructs made using traditional cloning methods (Table 2.2.1).

Table 2.2.1. Details of constructs made using traditional cloning methods.

Protein	Construct	Tag	Organism	Vector
TBL1X	1-527	N-His_{6x}	M. musculus	pFL
TBL1X	148-527	N-His_{6x}	M. musculus	pFL
TBL1X	176-527	N-His_{6x}	M. musculus	pFL
TBL1XR1	1-514	N-His_{6x}	M. musculus	pFL
TBL1XR1	134-514	N-His_{6x}	M. musculus	pFL
TBL1XR1	163-514	N-His_{6x}	M. musculus	pFL
TBL1XR1	1-514	N-His_{6x}	M. musculus	pmCherryN1

2.2.1 Insert amplification using PCR

All the inserts for cloning were generated via PCR from either cDNA or plasmid containing the cDNA. PCR cycling was carried out using PfuUltra II HS DNA polymerase from Agilent Technologies following a manufacturer-specified PCR protocol. The PCR reaction buffer composition was altered by adding DMSO to a final concentration of 3% to inhibit DNA secondary structure formation. Furthermore, 0.5 µl of the DNA polymerase, instead of 1 µl, was used. Finally, instead of 0.5 µl of 25 mM dNTPs, 1 µl of 10 mM dNTPs were used (Table 2.2.2).

Table 2.2.2. Standard PCR reaction buffer composition.

Component	**Amount**
DNA template	50-100 ng
Primers (10 µM)	1 µl (each)
dNTPs (10 mM)	1 µl
DMSO (100%)	1.5 µl
PfuUltra II reaction buffer (10X)	5 µl (1X)
PfuUltra II HS DNA polymerase	0.5 µl
H_2O	To a volume of 50 µl

2.2.2 Restriction digest

In general, approximately 1 µg of DNA was added to a solution containing 10 units of appropriate restriction enzymes (all from New England Biolabs, NEB) to a total volume of 20 µl (see Table 2.2.3 for details). The mixture was incubated for one hour at temperatures appropriate to the enzymes used, after which they were either heat-inactivated according to manufacturer's instructions and used directly, or if heat-inactivation was not possible, the reaction mix was separated on a 1% (w/v) agarose gel using SYBR® Safe (Thermo Fisher) as DNA stain. The corresponding DNA band was cut out and purified using QIAGEN QIAquick Gel Extraction kit and eluted in 30 µl of ddH_2O.

Table 2.2.3. Standard restriction digest reaction.

Component	Final amount
Restriction enzyme(s)	10 Units (each)
DNA	1 µg
Buffer (10X)	2 µl (1X)
H_2O	To 20 µl

2.2.3 Ligation

Both vector and insert were processed the same way (section 2.2.2), after which they were ligated together (Table 2.2.4). The reaction consisted of 100 ng of vector DNA and 200-500 ng of insert DNA depending on the size of the insert. T4 DNA ligase (from NEB) and its buffer were added to the mix and the whole solution made up to 20 µl with ddH_2O. This was allowed to ligate at room temperature for an hour. Alternatively, due to time-constraints the incubation was sometimes carried out overnight at 4˚C.

Table 2.2.4. Standard ligation reaction mix components.

Component	Final amount
T4 DNA ligase	400 Units
T4 DNA ligase buffer (10X)	2 µl (1X)
Vector DNA	100 ng
Insert DNA	200-500 ng
H_2O	To 20 µl

The ligated vector was used for transformation with either electro- or chemically competent cells, depending on use and availability, the protocols of which are listed in sections 2.2.4.1 and 2.2.4.2.

2.2.4 Transformation

2.2.4.1 Electrocompetent E. coli

A 50 µl aliquot of electrocompetent *E. coli* cells was thawed on ice. Approximately 100 ng of plasmid DNA in a volume of 1-2 µl was added to the thawed cells. The tube was flicked a few times to mix the cells with DNA after which the contents were transferred to an ice-cold

MicroPulser™ Electroporation Cuvette (0.2 mm gap width) (Bio-Rad) and pulsed with a MicroPulser™ Electroporator using the pre-set Ec2 protocol. 200 µl of room temperature SOC was immediately added to the cells. The cells were grown for 1 hour at 37˚C with shaking at 850 RPM in an Eppendorf Thermomixer® Compact. All the cells were plated out onto a LB agar plate containing appropriate antibiotics and grown overnight at 37˚C.

2.2.4.2 Chemically competent E. coli

A 50 µl vial of chemically competent E. *coli* cells was thawed on ice. Approximately 100 ng of plasmid DNA in a volume of 1-2 µl was added to the thawed cells. This was allowed to incubate on ice for 30 minutes, after which the cells were subjected to a heat shock at 42˚C for 30 seconds, unless specified otherwise. After the heat shock, the cells were allowed to recover on ice for 3 minutes, after which 200 µl of SOC was added and the cells grown at 37˚C for 1 hour. All the cells were plated onto an LB agar plate with appropriate antibiotics and grown overnight at 37˚C.

2.2.5 Plasmid amplification and verification

From an overnight growth on a plate, between 8 and 12 individual colonies were inoculated into tubes containing 5 ml of LB medium with appropriate antibiotics and grown overnight with shaking (200 RPM) at 37˚C. The plasmid DNA was purified according to manufacturer's specifications using QIAGEN QIAprep Spin Miniprep kit and eluted with 30 µl of ddH_2O. DNA concentration was measured using a NanoDrop 2000 (Thermo Scientific) at 280 nm. Plasmids were sequenced using the BigDye® reaction by Edinburgh Genomics.

2.3 Ligation-independent cloning

The following constructs were made using ligation-independent cloning (LIC) (Table 2.3.1).

Table 2.3.1. Constructs made using Ligation-Independent Cloning.

Protein	Construct	Tag	Organism	Vector
TBL1X	1-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	127-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	139-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	150-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	162-527	N-His_{6x}	*M. musculus*	pEC-KHC

TBL1X	171-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	176-527	N-His_{6x}	*M. musculus*	pEC-KHC
TBL1X	176-527 Codon optimised	N-His_{6x}	*M. musculus*	pEC-KHC

For LIC, both vector and insert were processed to have specific complementary overhangs which are created due to the T4 DNA polymerase 3' to 5' exonuclease activity. Normally the exonuclease activity is balanced by the polymerase activity, however by only adding one of the four dNTPs, it is possible to make the polymerase stall at a specific location. This is what creates the complementary overhangs (Figure 2.3.1).

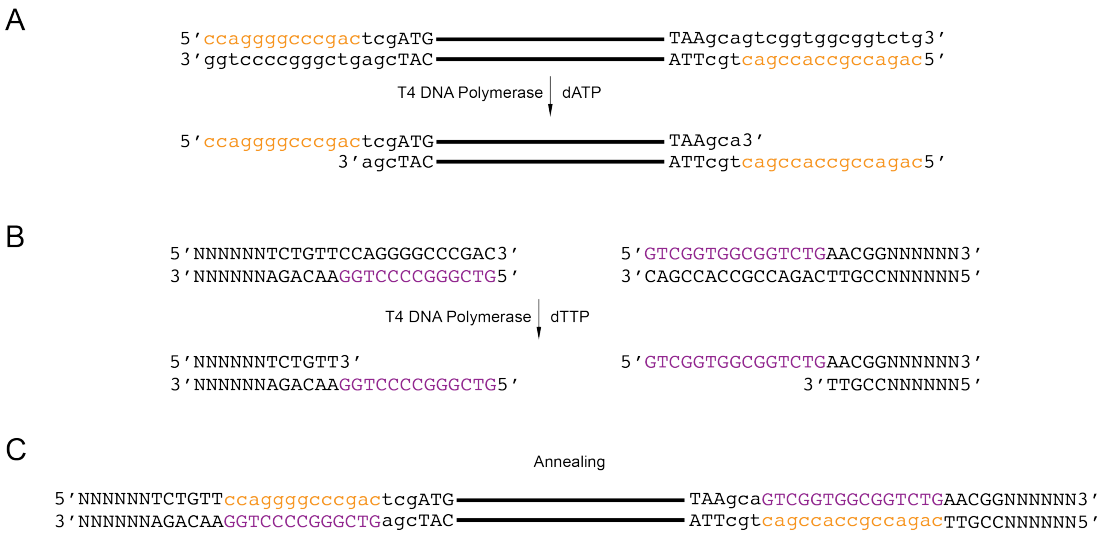

Figure 2.3.1. Principle of ligation-independent cloning. A) Gene of interest (black line) is amplified using PCR with primer extensions shown in black and orange. The amplified fragment is processed with T4 DNA Polymerase which has has 3' to 5' exonuclease activity. This is done in the presence of only dATP, which creates the overhangs. B) Vector is cut with an enzyme that creates the two blunt ends seen. This is processed with T4 DNA polymerase in the presence of dTTP, creating the overhangs (purple). C) The processing generates complementary overhangs that anneal together. This DNA can be used directly in transformation without the need to ligate the fragments together.

For vector processing, 2 µg of vector was digested with 20 U of ZraI in manufacturer specified buffer. 250 ng of linearised vector was loaded into each well in a 0.8% agarose gel. The linearised vector bands were cut out form the gel and purified using a QIAquick Gel Extraction Kit (Qiagen) following manufacturer's instructions. 450 ng of linearised vector was mixed with 3 µl of T4 DNA polymerase buffer (10X), 3 µl of dTTP (25 mM), 1.5 µl of DTT (100 mM) and 0.6 µl of T4 DNA polymerase (LIC qualified, Novagen). The volume was made up to 30 µl

using ddH_2O. The mixture was incubated at room temperature for 30 minutes, followed by heat inactivation of the enzyme at 75˚C for 20 minutes.

The insert was processed as follows: 600 ng of gel-purified PCR product was mixed with 2 µl of T4 DNA Polymerase buffer (10X), 2 µl of dATP (25 mM), 1 µl DTT (100 mM) and 0.4 µl of T4 DNA Polymerase (LIC qualified, Novagen) was added. The volume was made up to 20 µl using ddH_2O. The mixture was incubated at room temperature for 30 minutes, followed by heat inactivation of the enzyme at 75˚C for 20 minutes.

The annealing reaction was set up by adding 2 µl of processed insert to 1 µl of processed vector. The mixture was left to incubate at room temperature for 10 minutes, followed by addition of 1 µl of 25 mM EDTA. This was incubated at room temperature for an additional 10 minutes, after which 2 µl of the sample was used for transformation into electro- or chemically competent *E. coli* cells.

2.4 Gibson Assembly

Gibson assembly was performed using the Gibson Assembly® Cloning Kit from NEB following the manufacturer's protocol. Sequence overlap of the primers was generally between 20 and 40 base pairs, although priority was given to both forward and reverse primers having similar melting temperatures rather than matching the primer lengths as suggested by the manufacturer.

Table 2.4.1. Constructs made using Gibson Assembly

Protein	Construct	Tag	Organism	Vector
TBL1XR1	1-514	C-His_{6x}	*M. musculus*	pEC-MCN-CS
TBL1XR1	127-514	C-His_{6x}	*M. musculus*	pEC-MCN-CS
TBL1XR1	159-514	C-His_{6x}	*M. musculus*	pEC-MCN-CS
TBL1XR1	163-514	C-His_{6x}	*M. musculus*	pEC-MCN-CS
TBL1XR1	1-514	No tag	*M. musculus*	pEC-MCN-CS
TBL1XR1	1-514	N-His_{6x}	*M. musculus*	pEC-MCN-CS
HDAC3	1-428	No tag	*H. sapiens*	pEC-MCN-CS
HDAC3	1-428	N-GST	*H. sapiens*	pEC-EAVNG

2.5 Site-directed mutagenesis

Table 2.5.1 contains a list of constructs made using site-directed mutagenesis. Complementary primer pairs (forward and reverse) with the desired mutation(s) were used to amplify the vector containing the gene of interest by PCR. DpnI (0.4 U/µl) was added directly to the PCR mix and incubated for 1 hour at 37˚C to remove the methylated template. 2 µl of the digested reaction mix was transformed into electro- or chemically competent *E. coli* cells.

Table 2.5.1. Constructs made using site-directed mutagenesis.

Protein	Construct	Tag	Organism	Vector
TBL1XR1	134-514 E171A	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 E171Q	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 C214S	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 D313N	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 E351A	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 E351D	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 D369A	N-His_{6x}	*M. musculus*	pFL
TBL1XR1	134-514 Y446F	N-His_{6x}	*M. musculus*	pFL

2.6 Insect cell cloning

2.6.1 Making EMBacY selective plates

The LB agar was melted in a microwave and left to cool (~50˚C). Antibiotics (Table 2.6.1) were added to the agar and plates poured and stored at 4˚C.

Table 2.6.1. Antibiotics used in making EMBacY selective plates

Antibiotic	Stock solution	Per 100ml
Kanamycin	50 mg/ml	100 µl
Tetracycline	10 mg/ml	100 µl
Carbenacillin	100 mg/ml	100 µl
Gentamycin	10 mg/ml	100 µl

On the day of use, 100 µl of IPTG (0.1 M stock) and 100 µl of X-gal (20 mg/ml stock in DMSO) were spread on each plate and left on the bench for five minutes to dry.

2.6.2 Making EMBacY competent cells

On day one, a vial of EMBacY cells was taken from the -80˚C freezer and a sterile loop was used to streak the cells onto an EMBacY selective plate (see 2.6.1). The cells were incubated at 37˚C for 16-20 h.

On day two, a single EMBacY colony was inoculated into 5ml of LB with appropriate antibiotics. The cells were grown overnight at 37˚C with shaking for 16-20 h.

On day three, the overnight culture was diluted in a ratio of 1:200 into 250 ml of pre-warmed LB medium (37˚C) containing 20 mM $MgSO_4$. Cultures were grown until they reached an OD_{600} of ~0.6. The 250 ml culture was split into two sterile chilled 250 ml centrifuge bottle and left to cool on ice for 10 minutes. The cells were pelleted by centrifugation at 4000 *g* at 4˚C for 5 minutes.

All of the supernatant was decanted and 50 ml of ice-cold TFB1 (Table 2.6.2) was added to each bottle. Cells were resuspended by gently pipetting them up and down, followed by a 5 minute incubation on ice. Cells were pelleted by centrifugation at 1400 *g* at 4˚C for 10 minutes.

The supernatant was discarded and cells were gently resuspended in 5 ml of ice-cold TFB2 (Table 2.6.2), followed by incubation on ice for 15 minutes. 100 µl aliquots were dispensed in a sterile environment into pre-chilled 1.5 ml eppendorfs and flash-frozen in liquid N_2. Cells were stored at -80˚C.

Table 2.6.2. Compositions of buffers used in making chemically competent EMBacY cells.

TFB1	TFB2
30 mM CH_3COOK	10 mM MOPS
100 mM RbCl	10 mM RbCl
10 mM $CaCl_2$	10 mM $CaCl_2$
50 mM $MnCl_2$	15% glycerol
15% glycerol	

2.6.3 EMBacY transformation

A 100 µl vial of EMBacY cells was thawed on ice. Approximately 40 ng of the transfer plasmid containing the gene of interest was added to the cells and tube flicked a few times to mix the DNA. The cells were left to sit on ice for 45 minutes after which they were heat shocked for 60 seconds at 42˚C. The cells were allowed to recover on ice for 3 minutes, followed by the addition of 900 µl of SOC and outgrowth at 37˚C with shaking for 3-4 hours.

The cells were spun down at 10,000 *g* at 4˚C for 1 minute. 750 µl of SOC was removed and discarded and the cells were resuspended in the remaining 250 µl of SOC. The cells were plated onto EMBacY plates that contained X-gal and IPTG (see section 2.6.1).

After 24-48 hours the plates were inspected. Between 5 and 10 white colonies, which were the result of recombination of the transfer plasmid into the middle of the LacZ gene and subsequent inability to metabolise X-gal, were picked and restreaked onto an EMBacY selective plate. After 24-48 hours white colonies were taken further for bacmid isolation.

2.6.4 Bacmid isolation

True positives from re-streaked EMBacY selective plates were picked and grown overnight in 5 ml of LB with appropriate antibiotics (Table 2.6.1) at 37˚C with shaking. The next day the cells were pelleted by centrifugation at 4000 *g* at 4˚C for 10 minutes and the supernatant discarded. The cells were resuspended (on ice) in 250 µl of buffer P1 from QIAGEN QIAprep Spin Miniprep kit. To this 250 µl of buffer P2 from the same kit was added, the tubes were inverted a few times to mix the contents and 350 µl of buffer P3 from the kit was added to stop the reaction. The resulting mix was centrifuged for 10 minutes at 10,000 *g* at 4˚C.

The supernatant was transferred to a new 1.5 ml Eppendorf while taking care to avoid the white sediment (cell debris). 800 µl of 100% propan-2-ol was added to the supernatant and the bacterial artificial chromosome (bacmid) DNA was pelleted by centrifugation at 10,000 *g* at 4˚C for 15 minutes. The supernatant was removed, taking care not to dislodge the DNA pellet. The pellet was washed with 500 µl of ice-cold 70% ethanol to remove excess salt. The solution was centrifuged for 5 minutes at 10,000 *g* at 4˚C, washed with ethanol, and centrifuged again.

All the ethanol was removed and the pellet was allowed to air dry. The DNA pellet was resuspended in 50-100 µl of ddH_2O. The concentration of the bacmid was then measured using a NanoDrop 2000 (Thermo Scientific) and diluted down to 1 µg/µl.

2.7 Generation of baculovirus stocks

2.7.1 Transfection of Sf9 cells

Transfection of Sf9 cells was carried out in 6-well plates in a sterile tissue culture hood. This generated the V0 virus stock that could be used to infect cells and generate more virus with higher titre. First, Sf9 cells were collected from a T75 flask that was 95-100% confluent. The old medium was removed and three times as much fresh medium (Sf900-III SFM, Gibco) was added. Cells were resuspended by gently hitting the flask against a hard surface.

To each well of a 6-well plate (Corning™ Costar™ Flat Bottom Cell Culture Plate) 2 ml of the suspended cells were added (final concentration $2*10^6$ cells/ml). The cells were placed in a 27˚C incubator for 30 minutes. In the meantime transfection mix (Table 2.7.1) and negative control (Table 2.7.2) were prepared. The Sf-900 III medium (Life Technologies) was pre-warmed to 27˚C prior to use. The X-treme GENE HP (Roche) transfection reaction was added directly into the mixture to minimize contact with plastic surfaces. The mix was left at room temperature for 30 minutes.

Table 2.7.1. Transfection mix for transfecting Sf9 cells.

Component	Amount
Bacmid DNA (1µg/µl)	40 µl
Sf-900 III medium	400 µl
X-treme GENE HP	16 µl

Table 2.7.2. Negative control for transfecting Sf9 cells

Component	Amount
Bacmid DNA (1µg/µl)	9 µl
Sf-900 III medium	90 µl

After thirty minutes, 90 µl of the transfection mix was added dropwise to each well (five in total), along with 90 µl of the negative control to sixth well. The plate was gently rocked to mix the contents and put into a 27˚C incubator. Once 90% of the cells reached fluorescence (usually after 72-96h), the medium, which contained the V0 virus stock, was collected from wells 1-5. The virus and cell mix was spun in a centrifuge at 400 *g* at 27˚C and the resulting

supernatant filtered through a 0.22 µm filter into a sterile tube. The virus stock (V0) was wrapped in foil to protect it from the light and stored at 4˚C.

2.7.2 Baculovirus amplification

The V0 stock from initial transfection was used to generate V1 stock that was then used to generate subsequent V2 stocks. All of the V0 stock (10 ml) was added to 200 ml Sf9 cells (1:20 virus:medium ratio) that were at a density of $2*10^6$ cells/ml. These cells were grown for 3 days or until >90% of the cells were fluorescent, whichever came first. The cells were pelleted by centrifugation at 500 *g* for 5 minutes at 27 ˚C, and the supernatant filtered through 0.22 µm filter to generate the V1 stock. The cell pellet was flash frozen in liquid N_2 and stored at -80˚C. The V1 stock was then used to generate V2 stock, using conditions identical to those used to generate V1. Again, both V1 and V2 virus stocks were wrapped in foil and stored at 4˚C. The cells from the generation of V1 and V2 stocks were stored at -80 ˚C.

2.8 Protein expression tests

2.8.1 Heterologous expression in *E. coli*

Initial screenings were done using *E. coli* in one or all of the following strains: BL21(DE3), BL21(DE3) RIPL, B834(DE3), T7 Express. The cells were transformed and plated according to the protocol in section 2.2.4. After overnight growth some cells were scraped off the plate and inoculated into 25 ml of LB. They were then grown at 37˚C until OD_{600} reached 0.6, after which a 1 ml uninduced sample was taken, spun down (16,000 *g* at 4˚C for 10 minutes), and resuspended in 50 µl of 2X sample buffer (2XSB, Tris HCl pH 6.8 100 mM, 10% ß-mercaptoethanol, 4% SDS, 0.2% bromophenol blue, 20% glycerol). The cells were induced with 1 mM IPTG and temperature set to 18˚C. The cells were grown overnight. In the morning a 1 ml induced sample was taken. The OD of the sample was measured, the cells were pelleted, and resuspended in 50 µl of 2XSB per unit of OD_{600}.

The rest of the cells were harvested (4500 *g* at 4˚C for 10 minutes) and resuspended in 5 ml of appropriate lysis buffer (Table 2.8.1) and sonicated 12 times on ice with an amplitude of 10 µm (10 seconds on / 5 seconds off). 2 ml of the lysed cells were transferred to a 2 ml Eppendorf tube and spun down at 16,000 *g* at 4˚C for 10 minutes.

In the meantime, 30 µl of appropriate slurry (50% v/v of resin in 20% ethanol) was washed with 1 ml of ddH_2O, followed by three washes with 1 ml of lysis buffer (pelleted between washes at 700 *g* / 4˚C / 1 min).

A 10 µl sample of the cleared supernatant (soluble fraction) was removed and mixed with 10 µl of 2XSB. The rest of the soluble fraction was added to the slurry, while the pellet was used as the insoluble fraction in subsequent analysis and resuspended in 200 µl of 2XSB. The slurry mix was put on a rocking platform in the cold room (4˚C) and left for 15-30 minutes. It was subsequently washed 3 times (700 *g* at 4˚C for 1 minute), after which the beads were resuspended in 15 µl of 2X loading buffer. 2 µl of uninduced, induced, soluble and insoluble, as well as 10 µl of pulldown fractions were loaded and analysed on an SDS-PAGE gel.

Table 2.8.1. Lysis buffers for small-scale pulldowns using Ni^{2+}-NTA or GST beads

Ni^{2+}-NTA	**GST**
20 mM Tris HCl pH 7.5	20 mM Tris HCl pH 7.5
150 mM NaCl	150 mM NaCl
0.5 mM ß-mercaptoethanol	1 mM DTT

If the expression tests showed no (soluble) protein production, induction temperatures and lengths were varied to determine whether different growth conditions would support (improved) protein expression.

2.8.2 Heterologous expression in Insect cells

Small-scale expression tests on Sf9 cells were carried out on the cells obtained from the generation of V2 stocks (see section 2.7.2). Hi5 cells for pulldowns were obtained in similar way by infecting the cells with V1 stock at a ratio of 1:20 virus:medium. The difference was that dextran sulphate was added to the Hi5 cells at a concentration of 100 mg per 1 L of culture to keep the cells from clumping together. Furthermore, because the medium lacked L-glutamine, 90 ml/L of medium of 200 mM L-glutamine was added. The cells were resuspended in appropriate lysis buffer (Table 2.8.1) and sonicated on ice at 10 µm amplitude (Soniprep 150, MSE) for 3 minutes at 30 sec on / 15 sec off. The rest of the small-scale pulldown protocol was identical to that in section 2.8.1.

2.9 MeCP2 "minigene" protein expression and purification

2.9.1 Large-scale expression

Bacterial cells were transformed and plated according to the protocol in section 2.2.4. After overnight growth, some cells were scraped off the plate and transferred into 50 ml of LB that contained appropriate antibiotics to serve as the starter culture. The cells were grown for an hour to allow them to recover and 5 ml of the starter culture was added into 250 ml of pre-warmed (37˚C) 2XTY medium containing appropriate antibiotics in 2 L flasks. The cells were grown at 37˚C with shaking until OD_{600} reached 0.6. The temperature was set to 30˚C and the cells were grown for 3 h before harvesting by centrifugation for 10 minutes at 4˚C and 4000 *g*. The pellets were weighed and flash-frozen in liquid N_2, after which they were stored at -80˚C.

2.9.2 Protein purification

Bacterial cell pellets were resuspended to a final concentration of 10% (w/v) in lysis buffer containing 20 mM Tris HCl pH 7.5, 500 mM NaCl, 10 mM imidazole, 0.5 mM β-mercaptoethanol, one complete EDTA free protease inhibitor cocktail tablet (Roche), and 1 mM Pefabloc (Roche). The cells were lysed on ice using sonication at 10 µm amplitude for 5 minutes at 30 sec on / 15 sec off. The lysate was spun down at 50,000 *g* at 4˚C for 45 minutes. The cleared lysate was added to 1 ml (per 20 g of wet cell mass) of equilibrated Ni^{2+}-NTA resin and incubated for 1 hour, after which the resin was spun down at 700 *g* for 5 minutes at 4˚C. If the protein was injected onto a cation exchange column the following day, the beads were washed three times (instant wash) with 50 ml of lysis buffer. The protein was eluted from the resin by washing it five times (instant wash) with 5 ml buffer (20 mM Tris HCl pH 7.5, 50 mM NaCl, 500 mM imidazole, 0.5 mM β-mercaptoethanol). Each fraction was analysed on an SDS-PAGE gel, followed by pooling of relevant fractions. The protein was dialysed overnight in Spectra/Por 6-8 kDa MWCO RC dialysis tubing (SpectrumLabs) into cation-exchange buffer (20 mM Tris HCl 7.5, 50 mM NaCl, 1 mM DTT). The following morning the dialysed protein sample was filtered through 0.22 µm filter and injected onto a 1 ml Resource S column (GE Life Sciences).

If the protein was applied to the cation exchange column the same day, the beads were washed three times with 50 ml of wash buffer (20 mM Tris HCl pH 7.5, 50 mM NaCl, 10 mM imidazole, 0.5 mM β-mercaptoethanol). The protein was eluted off the resin by washing it

five times of 5 ml each (20 mM Tris HCl pH 7.5, 50 mM NaCl, 500 mM imidazole, 0.5 mM β-mercaptoethanol) and analysed on an SDS-PAGE gel, followed by pooling of relevant fractions. The protein was filtered through 0.22 µm filter and injected onto a 1 ml Resource S column (GE Life Sciences).

In either case, after injecting the protein onto the cation-exchange column, the column was washed to baseline absorbance with buffer A (20 mM Tris HCl pH 7.5, 50 mM NaCl, 1 mM DTT). After reaching baseline, the column was washed with 30% buffer B (20 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT) and washed to baseline again. MeCP2 minigene protein was eluted with a gradient of 30-100% buffer B over 30 column volumes, collecting 1 ml fractions. Fractions were analysed and pooled, followed by concentrating the sample down to approximately 2.5 ml for injection onto a size-exclusion chromatography column (Superdex 75 16/600 prep grade) equilibrated with 20 mM Tris HCl pH 7.5, 200 mM NaCl, 1 mM DTT. The resulting fractions were analysed on an SDS-PAGE gel, pooled, and concentrated using VivaSpin 6 10 kDa MWCO concentrator. The protein was stored at -20˚C as 10 µl aliquots flash frozen in liquid N_2.

2.10 Expression and purification of TBL1X, TBL1XR1, and TBL1XR1 mutants

2.10.1 Large-scale expression

Sf9 cells were infected at a density of $2*10^6$ cells/ml with V2 virus at a ratio of 1:20 to 1:50 virus:medium, depending on the strength of the virus stock. Antibiotic-antimycotic (100X solution, Thermo Scientific, catalogue number 15240-096) solution was added to a final concentration of 1X. These cells were allowed to grow for three days after which they were harvested by centrifugation at 3000 *g* for 10 minutes at 4˚C. Cell fluorescence and health was monitored, after which adjustments could be made to the amount of virus used, if needed. The pellets were weighed and flash-frozen in liquid N_2, after which they were stored at -80˚C.

2.10.2 Protein purification

Insect cell pellets were resuspended to a final concentration of 20% (w/v) in lysis buffer containing 20 mM Tris HCl pH 7.5, 500 mM NaCl, 10 mM imidazole, one complete EDTA free protease inhibitor cocktail tablet (Roche), and 1 mM Pefabloc. The cells were lysed on ice

using sonication at 10 µm amplitude for 3 minutes at 30 sec on / 15 sec off. The lysate was spun down at 50,000 *g* at 4˚C for 45 minutes.

The lysate was added to 1 ml of pre-equilibrated NiNTA resin (equilibrated with lysis buffer) per 20 g of wet cell pellet in batch and incubated at 4˚C with rolling for 1 hour. The resin was spun down at 700 *g* for 5 minutes at 4˚C, after which the supernatant was discarded and the beads washed three times with 50 ml of wash buffer containing 20 mM Tris HCl pH 7.5, 500 mM NaCl, 10 mM imidazole. The protein was eluted with 5x5 ml of 20 mM Tris HCl pH 7.5, 500 mM NaCl, 500 mM imidazole. Eluted protein was transferred into a Spectra/Por 6-8 kDa MWCO RC dialysis tubing (SpectrumLabs) and dialysed overnight at 4˚C into 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT.

The following day the dialysed sample was filtered using 0.22 µm filters (Millex Millipore) and concentrated using VivaSpin R15 30 kDa MWCO concentrators (Sartorius) to a volume of approximately 2.5 ml. This sample was subjected to size-exclusion chromatography using HiLoad Superdex 16/600 75 prep grade (GE Healthcare Life Sciences) equilibrated in the dialysis buffer. The fractions were analysed on an SDS-PAGE gel and appropriate fractions were pooled and concentrated using VivaSpin 6 10 kDa MWCO concentrator. The protein was stored at -20˚C as 10 µl aliquots flash frozen in liquid N_2.

2.11 Co-crystallisation of TBL1X/TBL1XR1 with MeCP2 NID peptide

2.11.1 Crystallisation screening

Initial crystallisation trials were carried out in 96-well MRC plates (Molecular Dimensions) using in-house Crystal Gryphon (ARI) nanolitre-dispensing robot. 0.1 µl of protein solution plus 0.1 µl of well solution was used for setting up the drops with 65 µl in the reservoir. In the initial scouting phase of the crystallisation trials, $TBL1XR1_{134\text{-}514}$ concentration was varied between 4 and 20 mg/ml. The crystallisation screens used were MAGIC I and MAGIC II (Crystallisation Facility, MPI Martinsried), Index HT (Hampton Research), JCSG+ (Molecular Dimensions), COMPLEX I and COMPLEX II (Crystallisation Facility, MPI for Biochemistry, Martinsried), Anions suite (Qiagen NeXtal), and AmSO4 suite (Qiagen NeXtal). The plates were sealed manually using polyolefin film (fluidX) and left at 18˚C. In a few cases, plates were set up at room temperature and transferred to 4˚C.

Hanging drops were set up in 24-well Linbro plates using 500 µl of well solution and a drop containing 1 µl protein solution mixed with 1 µl of well solution. The protein solution contained varying concentrations of $TBL1XR1_{134-514}$ mixed with $MeCP2_{285-309}$ in 1:1.2 ratio that had been incubated at 4˚C for an hour prior to setting up the drops. Conditions were varied to determine the effect PEGs of different molecular weight, or PEG concentration, although a few screens also examined the effect of pH on crystallisation. Different protein concentrations were screened, as well a different protein:well solution ratios. The best diffracting crystals were obtained in 100 mM MOPS pH 7.5, 18% PEG 3350 using 10 mg/ml of $TBL1XR1_{134-514}$ (final concentration) mixed with $MeCP2_{285-309}$ in 1:1.2 ratio.

2.11.2 Data collection

Crystals were harvested by transferring them into cryoprotectant solution containing the crystallisation condition buffer supplemented with 30% v/v of glycerol and flash-cooled in liquid nitrogen. The crystals were shipped to Diamond Light Source (DLS) in Oxford. The data from the best diffracting crystal was collected on the i03 beamline using a strategy calculated automatically by MOSFLM (Battye et al., 2011). A total of 1200 images were collected using a rotation of 0.20˚ (total rotation 240˚), 0.2 second exposure and a wavelength of 0.9763 Å.

2.11.3 Data processing

XDS was used to index and integrate reflections from all the images collected (Kabsch, 2010). The majority of settings for the XDS input file were taken from the Diamond Light Source data file, however Friedel's Law option was set to TRUE because there was no anomalous scattering present. The resolution range used was 50 Å – 2.5 Å.

Next, the file with all the corrected spot intensities was used as the input for COMBAT from the CCP4 Suite (Winn et al., 2011). This converted the .HKL file into an .mtz file which allowed to use SCALA to scale and merge each observed intensity into an average value.

2.11.4 Molecular replacement

A structure of the human TBL1XR1 (PDB ID: 4LG9) was used for molecular replacement. A sequence alignment file (*.aln) using the human TBL1XR1 and mouse TBL1XR1 was also created. The model used for molecular replacement was altered by deleting all unknown atoms and ions in the structure (labelled UNX), as well as all water molecules. Furthermore, loops that are variable in sequence between the two proteins were removed. The sequence alignment and the PDB file were used as an input for CHAINSAW (CCP4), which was used to

trim down all non-conserved residues to the C-alpha atom in the model PDB file. The output from CHAINSAW was used as template for PHASER (McCoy et al., 2007). Additionally, based on the calculated Matthew's coefficients (V_M) for different number of molecules in the asymmetric unit (Equation 2.1 and 2.2), the most likely asymmetric unit contained 4 molecules with a solvent content of 44% (Table 2.11.1), as the mode of Matthew's coefficient is 2.34 Å/Da (Kantardjieff and Rupp, 2003). This was also used as an input parameter for PHASER.

Table 2.11.1. Calculated Matthew's coefficients and solvent content for different number of molecules in the asymmetric unit.

Molecules in the asymmetric unit	Matthew's coefficient ($Å^3$/Da)	Solvent content (%)
1	8.87	86.14
2	4.43	72.28
3	2.96	58.41
4	2.22	44.55
5	1.77	30.69
6	1.48	16.83
7	1.27	2.97

$$V_M = \frac{a*b*c}{n*M_w} \text{ (Eq 2.1)}$$

$$\%_{solvent} = 100 - \frac{1.23}{V_M} * 100 \text{ (Eq 2.2)}$$

The validity of the PHASER solutions was assessed based on the Z-score and the log likelihood gain (LLG) values. According to the PHASER manual, solutions with a translation function Z-score (TZF) of more than 8 are classified as solved, and the LLG value should go up with each added component. The corresponding values are shown in Table 2.11.2. The building of the structure was done in COOT (Emsley et al., 2010), and refinement in Phenix (Adams et al., 2010). The structure was visualised in PyMol (Delano Scientific). The final structure was validated by looking at the r.m.s.d of bond angles and bond lengths and making sure there are no outliers, as well as examining reported Ramachandran outliers and clashes between atoms. Finally, both R_{work} and R_{free} were examined throughout the refinement to make sure the structure was not overinterpreted.

Table 2.11.2. PHASER TZF and LLG values for molecular replacement step. * - Origin is chosen randomly so no need for translation function

Molecule number	TFZ	LLG
1	N/A*	461
2	40.7	1790
3	13.8	333
4	13.0	504

2.12 Thermal Denaturation Assays

All thermal denaturation assays were done in triplicate in clear 96-well iQ Real-Time PCR plates (Bio-Rad) using iQ5 rtPCR Thermocycler (Bio-Rad) and 5X SYPRO Orange as a fluorescent dye (final concentration) unless stated otherwise. The final volume in each well was 50 µl. For all of the assays, the temperature range tested was 20-90˚C in 1˚C increments with a hold time of 30 seconds for each increment. Initially the protein was tested at different concentrations to see which concentration would give the best signal-to-noise ratio. After finding an appropriate concentration of protein to use, different additives and/or buffers were used to see how they affect the stability of the proteins. The melting temperatures were derived from the minimum of the negative first derivative of the measured fluorescence by plotting the change in the fluorescence (-d(RFU)/dT) against the temperature.

2.13 Surface Plasmon Resonance Assays

2.13.1 Wild-type TBL1XR1$_{134\text{-}514}$ binding to MeCP2 peptides

To determine the binding affinities and binding dynamics of TBL1XR1 binding to MeCP2$_{285\text{-}313}$ peptide, a surface plasmon resonance assay was used. All the experiments were performed at 25˚C using a BIAcore T200 (GE Life Sciences) and an SA chip (Series S Sensor Chip, GE Healthcare).

All 4 channels were first washed with buffer (20 mM Tris HCl pH 7.5, 200 mM NaCl, 1 mM DTT). Next 20 nM solutions of MeCP2 peptides were injected onto the chip to maximum Response Unit (RU) values of about 10 (R_{ligand}). This ensured that the maximum RU values (R_{max}) obtained upon TBL1XR1$_{134\text{-}514}$ binding were roughly 100, taking into account the molecular mass of the receptor ($Mr_{receptor}$), and the ligand (Mr_{ligand}) (Eq 2.1). Channel 1

was left as reference, while channel 2, 3, and 4 had wild-type MeCP2$_{285-313}$, wild-type MeCP2$_{285-309}$, and MeCP$_{285-313}$ R306C peptides immobilized on them, respectively.

$$R_{max} = R_{ligand} * \frac{Mr_{receptor}}{Mr_{ligand}} \quad (2.1)$$

30 µl/min of TBL1XR1$_{134-514}$ supplemented with 100 U/ml of heparin was injected over the channels for 30 seconds at various concentrations varying from 0.625 µM to 40 µM, followed by washing the chip for 180 seconds with buffer after each concentration. The data were analysed using the BIAcore T200 Evaluation software. All the affinities were derived from steady-state kinetics.

2.13.2 TBL1XR1$_{134-514}$ mutants binding to MeCP2 peptides

The effects of TBL1XR1 site directed mutants on binding MeCP2 were studied as described in section 2.13.1. All the protein samples were injected over the chip in succession without regeneration between runs of different mutants. All kinetic data were derived from steady-state kinetics and analysed using BIAcore T200 Evaluation software.

2.14 Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSAs) were used to determine whether MeCP2 minigene protein had a functional MBD and could discriminate between methylated and non-methylated DNA. They were also used to test whether binding of DNA to MeCP2 MBD influenced the affinity of the MeCP2 NID towards TBL1XR1 by generating a supershifted band.

2.14.1 DNA oligo annealing

The unlabelled and fluorescently-labelled (Fluorophore: DY-681) single-stranded DNA oligos (Biomers GmbH) (Table 2.14.1) were resuspended in nuclease-free water to a final concentration of 1 mM. 25 µl of labelled and 30 µl of unlabelled (1:1.2 molar ratio) DNA was mixed, giving a final concentration of 454 µM labelled and 545 µM unlabelled DNA. The mixture was heated up at 95˚C for 5 minutes in a heat block. After 5 minutes the heat block was wrapped in foil and left to cool down overnight, allowing the oligos to anneal slowly.

Table 2.14.1. Fluorescently-labelled methylated DNA used in EMSAs. The top strand had the 5' modification with DY-681 fluorophore, while the bottom stand was unlabeled. The '5' in the sequence stands for 5-methylcytosine.

Top strand (5' DY-681)	5' ata gaa gaa ttc 5gt tcc ag 3'
Bottom strand	3' at ctt ctt aag g5a agg tct 5'

2.14.2 DNA-binding assay

An 8% acrylamide/0.5X TBE gel was assembled and pre-run at constant power (2 W) for an hour at 4˚C. EMSA reactions were set up to a total volume of 20 µl, using either 0.5 µM or 0.1 µM DNA and varying the concentration of MeCP2 minigene protein. The buffer was 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA and 0.5 mM DTT. In some experiments, a dA/dT unlabelled competitor was added to try to distinguish between specific and non-specific binding. The amount of the competitor DNA varied during optimisation but was generally between 500 and 1500 ng per reaction (25 ng/µl to 75 ng/µl). Reactions were incubated at 4˚C in the dark for 15 to 30 minutes. 4 µl of native loading dye (0.25% w/v bromophenol blue, 50% v/v glycerol, 0.25% w/v xylene cyanole FF) was added to the samples and 8 µl of the sample was loaded onto the gel. The sample was separated on the gel at 4˚C with constant power (2 W) for approximately 90 minutes, after which the gel was imaged using Odyssey scanner (LICOR) using a wavelength of 700 nm.

2.14.3 TBL1XR1 supershift assay

TBL1XR1 supershift assays were performed as the DNA-binding assays in section 2.14.2 with the exception that the percentage of the polyacrylamide gel was lower (4%) and 40 µM $TBL1XR1_{134\text{-}514}$ was added to the reaction mix to a total volume of 20 µl.

2.15 Tryptic Peptide Identification by Mass-spectrometry

For mass spectrometry, samples were separated on an SDS-PAGE gel and the resulting band of interest cut out with minimal amount of extra acrylamide. The gel piece was incubated for 30 minutes at room temperature in 300 µl of 200 mM ammonium bicarbonate containing 50% acetonitrile. The supernatant was removed and the wash repeated. The gel pieces were then incubated for 1 hour at room temperature in 300 µl of 200 mM ammonium bicarbonate containing 50% acetonitrile and 20 mM DTT. The supernatant was removed and the gel piece

washed 3x (no incubation) with 300 µl of 200 mM ammonium bicarbonate in 50% acetonitrile.

The cysteines were alkylated by adding 100 µl of 50 mM iodoacetamide, 200 mM ammonium bicarbonate, and 50% acetonitrile to the gel. The tube was wrapped in foil and left in the dark for 20 minutes at room temperature. Again the supernatant was removed and the gel piece washed 3x (no incubation) with 300 µl of 200 mM ammonium bicarbonate in 50% acetonitrile.

The gel piece was covered with acetonitrile and left until the gel turned white. If the gel remained transparent, the supernatant was removed and fresh acetonitrile was added. The gel pieces were swollen in 30 µl of 50 mM ammonium bicarbonate containing 1 µl (2 µg/µl) of trypsin (Promega). The tube was sealed with parafilm and incubated overnight at 32˚C. The samples were stored at -80˚C until the analysis. On the day prior to analysis, the samples were sonicated for 5 minutes in a waterbath. Data were collected using a Brüker UltraflexExtreme MALDI TOF-TOF. Data were analysed using either MS-Fit from ProteinProspector specifying the protein sequence, mass tolerance of 70 ppm and constant carbamidomethyl modification, or using Mascot to perform Peptide Mass Fingerprinting with the same values and Swissprot database (Baker and Clauser, Mascot).

2.16 Fluorescence Polarisation Assays

All assays were done in black opaque 96-well plates (Greiner Bio One) in triplicate at room temperature with a final well volume of 100 µl using SpectraMax M5 multimode plate reader. Initially a concentration series of the fluorescent peptide diluted in same buffer as the protein involved in the binding experiment was used to determine the concentration of peptide needed (usually 10-200 nM) to get a good signal-to-noise ratio. That concentration of peptide was used to probe a range of concentrations with the protein of interest, both above and below the suspected K_D. If necessary, due to formation of aggregates or meniscus effects, the samples were supplemented with 0.005% Tween 20 (Thermo Fisher). Datapoints were analysed using non-linear regression with eq. 2.2 where R_f and R_b are the anisotropies of the free and bound receptor, respectively. L_T is the total ligand concentration, and K_D is the dissociation constant. Anisotropy is plotted as the Y value and X value plotted is protein concentration.

$$Y = R_f + (R_b - R_f) * \frac{\left(L_T + K_D + X - \sqrt{(-L_T - K_D - X)^2 - 4 * K_D * X}\right)}{2 * L_T} \quad \textbf{(2.2)}$$

2.17 Ethanol precipitation of DNA

To the DNA sample, 1/10th the original volume of 3 M CH_3COONa, pH 5.2, was added and mixed. 2.5 times the new volume of ice-cold 100% ethanol was added to the sample and again mixed. The DNA sample was placed into a -20˚C freezer for 30 minutes, after which the sample was centrifuged at 13,000 *g* for 15 minutes at 4˚C. The supernatant was removed and the DNA pellet washed with 1 ml of ice-cold 70% ethanol, spun down for 1 minute at 4˚C at 13,000 *g*, and the wash repeated one more time. All of the supernatant was subsequently removed and the DNA pellet air dried. The DNA pellet was resuspended in 30 µl of ddH_2O.

2.18 Pulldown assays of recombinant purified TBL1X with MeCP2

Pulldown assays were carried out using purified mouse $TBL1X_{176-527}$ and chemically synthesised N-biotinylated $MeCP2_{285-313}$ peptides containing K304E, K305R, K306C, and R309X mutations, as well as with phosphorylated T308 and wild-type $MeCP2_{285-313}$. Streptavidin (SA) beads were washed three times (1 ml wash) with NEI buffer (20 mM HEPES pH 7.5, 1 mM $MgCl_2$, 10 mM NaCl). 60 µl of the peptide was added to 20 µl of beads which were then incubated with rocking and rolling at 4°C for 1 h. NEI buffer was adjusted to contain 0.1% Triton-X-100, 1 mM DTT and 150 mM NaCl, which was used to wash the SA beads 4X (1 ml wash). Recombinant $TBL1X_{176-527}$ was added to the beads and incubated with rocking and rolling at 4°C for 1 h. Beads were washed 4X (1 ml wash) with NEI. Samples were boiled in 20 µl of 2X sample buffer for 10 min. 10 µl of the sample was loaded onto a 12.5% SDS-PAGE gel for analysis.

2.19 Pulldowns of TBL1XR1 from crude insect cell lysates with MeCP2

V1 cell pellets obtained from the generation of V2 virus stocks were resuspended in 5 ml of lysis buffer (20 mM Tris HCl pH 7.5, 150 mM NaCl, and 0.5 mM ß-mercaptoethanol) and lysed by sonication on ice at 10 µm amplitude (Soniprep 150, MSE) for 3 minutes at 30 sec on / 15 sec off. 60 µl of the peptide was added to 20 µl of beads which were then incubated with rocking and rolling at 4°C for 1 h. The beads were washed 4X (1 ml wash) in the lysis buffer.

1 ml of the crude lysate was added to the SA beads and incubated with rocking and rolling at 4°C for 1 h. Beads were washed 4X (1ml wash) with NEI. Samples were boiled in 20 µl of 2X sample buffer for 10 min. 10 µl of the sample was loaded onto a 12.5% SDS-PAGE gel for analysis.

3 Expression, purification and characterisation of TBL1X and TBL1XR1

3.1 Introduction

Previously, unpublished experiments done by Dr. Matthew Lyst and his summer student Žygimantė Tarnauskaitė in Prof. Adrian Bird's lab in Edinburgh showed that the minimal construct of TBL1X that interacts with MeCP2 contained residues 176-527, which corresponds to the C-terminal WD40 domain of TBL1X. Removing amino acids from the C-terminus ($TBL1X_{176-488}$) abolished the binding of MeCP2 to TBL1X in pulldown experiments, and constructs starting from amino acid 200 ($TBL1X_{200-527}$) also resulted in the loss of binding. This information was the rationale for the intial cloning of TBL1X and TBL1XR1 constructs.

To successfully complete this project, milligram quantities of recombinant protein were required at high purity. At the start of the project, a literature search revealed no previously established purification protocols for the WD40 domain of TBL1X or TBL1XR1. In this chapter, purification protocols for mouse TBL1X and TBL1XR1 constructs expressed in both *E. coli* and *S. frugiperda* cells were established. The purified proteins were characterised biophysically to determine their stability and permissible assay conditions for future experiments.

3.2 Construct design and cloning

Prior to the start of the project, two constructs of human TBL1X isoform 2, $TBL1X_{141-527}$ and $TBL1X_{159-527}$, had been previously cloned by Dr. Atlanta Cook. A number of mouse TBL1X and TBL1XR1 constructs were cloned into either bacterial or insect cell expression vectors (Table 3.2.1). All of the constructs except the full-length TBL1X and TBL1XR1 lacked the N-terminal LisH domain, which is the tetramerisation domain, but kept the C-terminal WD40 domain intact. The bacterial expression constructs were based on multiple secondary structure prediction softwares and care was taken not to disrupt any of these predicted structural elements as that could lead to insoluble protein. Furthermore, one of the constructs, $TBL1X_{176-527}$, was chosen for codon optimisation (synthesized by MWG) to determine whether this would enhance bacterial expression after initial expression tests showed promising results.

Table 3.2.1. Constructs of TBL1X and TBL1XR1 cloned for protein expression in *E. coli*. $TBL1X_{141-527}$ and $TBL1X_{159-527}$ were cloned by Dr. Atlanta Cook.

Protein	Construct	Tag	Organism	Vector
TBL1X	1-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	127-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	139-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	141-527	N-GST	H. sapiens	pEC-KGC
TBL1X	150-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	159-527	N-GST	H. sapiens	pEC-KGC
TBL1X	162-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	171-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	176-527	N-His_{6x}	M. musculus	pEC-KHC
TBL1X	176-527 Codon optimised	N-His_{6x}	M. musculus	pEC-KHC
TBL1XR1	1-514	C-His_{6x}	M. musculus	pEC-MCN-CS
TBL1XR1	127-514	C-His_{6x}	M. musculus	pEC-MCN-CS
TBL1XR1	159-514	C-His_{6x}	M. musculus	pEC-MCN-CS
TBL1XR1	163-514	C-His_{6x}	M. musculus	pEC-MCN-CS

3.3 Expression tests of TBL1X/TBL1XR1 in Escherichia coli

Initially expression tests for TBL1X and TBL1XR1 were done in *Escherichia coli* BL-21(DE3) cells. Pulldowns done by Dr. Uma Jayachandran on the two human TBL1X constructs ($hTBL1_{141-527}$ and $hTBL1_{159-527}$) showed that these two constructs expressed the protein of interest, though at low levels. It was decided to try to purify $TBL1X_{141-527}$ construct and establish a purification protocol, after which expression optimization could be undertaken. However, after multiple attempts of purification, $TBL1X_{141-527}$ always aggregated, as observed by size exclusion chromatography. Because of the low expression levels and aggregation problems with the $TBL1X_{141-527}$ construct, as well as the desire to perform biophysical assays on the mouse TBL1X/TBL1XR1 due to the availability of various mouse MeCP2 peptides in the Bird lab, these two human constructs were put aside. Instead, the expression levels of multiple mouse TBL1X constructs were tested. Two of them, $TBL1X_{162-527}$ and $TBL1X_{176-527}$, showed expression of the protein of interest, although the expression levels of these two constructs did not exceed that of $TBL1X_{141-527}$ and $TBL1X_{159-527}$ (Figure 3.3.1). At the same time, as most of the TBL1X constructs did not express any protein (data not shown), several TBL1XR1 constructs were cloned into different vector with a C-terminal His-tag to see if that

influenced protein expression. However none of these constructs showed any protein expression in BL21(DE3) cells (Table 3.3.1).

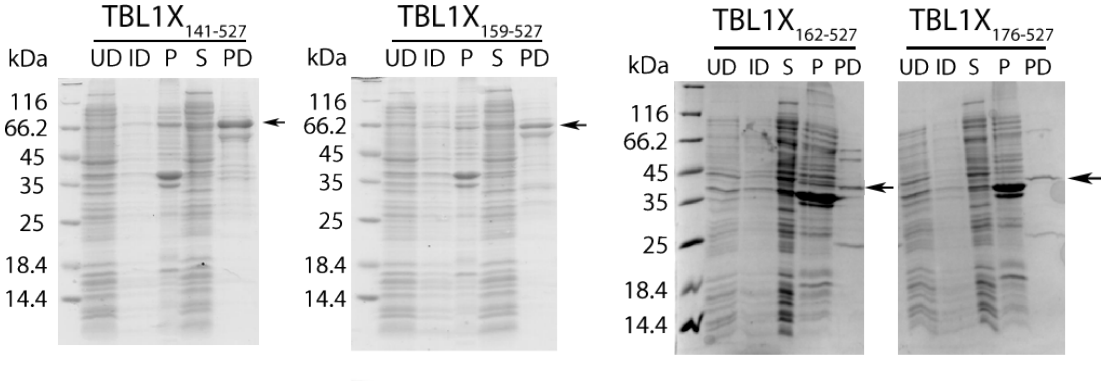

Figure 3.3.1. Expression tests of TBL1X constructs. Expression tests for $TBL1X_{141-527}$ and $TBL1X_{159-527}$ were carried out by Dr. Uma Jayachandran. Bands corresponding to the molecular weight of the proteins of interest are marked with a black arrow. The gel for $TBL1X_{162-527}$ and $TBL1X_{176-527}$ has been digitally cut in half to add the arrow but has not been shifted in any other way. UD – uninduced; ID – induced; S – soluble; P – pellet; PD – pulldown.

Table 3.3.1. All constructs of mouse TBL1X and TBL1XR1 which were tested for expression in BL-21(*DE3*) cells. Several of the clones did not express at all while most of those that did had minimal expression that was not sufficient for further characterisation. Both 162-527 and 176-527 TBL1X constructs expressed at quantities sufficient to warrant purification attempts. hTBL1X refers to human version of the protein.

Protein	Construct	Expresses	Soluble protein
TBL1X	1-527	No	
TBL1X	127-527	No	
TBL1X	139-527	No	
hTBL1X	141-527	Yes	Yes
TBL1X	150-527	No	
hTBL1X	159-527	Yes	Yes
TBL1X	162-527	Yes	Yes
TBL1X	171-527	No	
TBL1X	176-527	Yes	Yes
TBL1X	176-527 Codon optimised	No	
TBL1XR1	1-514	No	
TBL1XR1	127-514	No	
TBL1XR1	159-514	No	
TBL1XR1	163-514	No	

The effect of different *E. coli* strains on the protein expression levels was tested. BL21 strains are deficient in OmpT proteases, and *DE3* designates the presence of a genomically inserted λDE3 lysogen that encodes for T7 RNA polymerase under the control of a lac-promoter. B834 is a methionine auxotroph, which can be used to label proteins with selenomethionine. BL21(*DE3*) pRIPL strain carries a plasmid encoding extra copies for the tRNAs that are the most rare in *E. coli* and can be depleted during protein expression. T7 Express cells are derived from BL21 and contain the T7 RNA polymerase under the control of a lac operon in the *E. coli* genome with no prophage present in the cells. As the prophage can sometimes activate and cause cell lysis, this may yield more cell mass and consequently more expressed protein.

Several strains were tested on a select number of constructs (Table 3.3.2). These were chosen based on which constructs were required for future experiments. None of the strains tested produced any soluble protein for $TBL1X_{176-527}$ or $TBL1X_{162-527}$. However, both T7 Express and B834 expressed full-length TBL1XR1 although again at very low levels that would have been challenging for biophysical assays.

Table 3.3.2. All constructs of TBL1X and TBL1XR1 that were tested on various strains of *E. coli*.

Protein	Construct	Strain (*E. coli*)	Soluble protein
TBL1X	162-527	RIPL	No
TBL1X	176-527	RIPL	No
TBL1XR1	1-514	RIPL	No
TBL1X	162-527	T7 Express	No
TBL1X	176-527	T7 Express	No
TBL1XR1	1-514	T7 Express	Yes
TBL1X	176-527	B834	No
TBL1XR1	1-514	B834	Yes

$TBL1X_{176-527}$ was taken further for purification trials for two interconnected reasons. First, as mentioned previously, it was the minimal fragment that had been shown to interact with MeCP2. As the end goal was the crystallization of the WD40 domain of this protein, any extra amino acids could prove detrimental for crystal growth as they could be mobile and interfere with crystal packing. Secondly, none of the other mouse constructs expressed TBL1X or TBL1XR1 at higher levels compared to $TBL1X_{176-527}$.

3.4 $TBL1X_{176-527}$ purification from *E. coli*

The cells were lysed under pressure and the resulting lysate spun down to separate the insoluble proteins and other cellular debris. The soluble fraction was separated over an IMAC (Immobilised metal-affinity chromatography) column charged with Ni^{2+} ions to allow the His-tagged $TBL1X_{176-527}$ construct to bind, after which the column was washed to remove any unbound proteins. $TBL1X_{176-527}$ was eluted using a linear imidazole gradient (0-500 mM). A considerable amount of protein bound to the column (1A9-1B2) and eluted in one peak at a concentration of ~100 mM imidazole (20% B). Although the beads were washed to baseline absorbance, the elution fractions still had quite a lot of contaminants. This was most likely due to low amounts of recombinant tagged protein present in the sample, which allowed other proteins to bind. However there was a band corresponding to the molecular weight of $TBL1X_{176-527}$ present in the fractions (Figure 3.4.1, Figure 3.4.2). The A_{260}/A_{280} ratio could not be determined accurately as the sensor was saturated during elution, as observed from the chromatogram.

On SDS-PAGE, there is no obvious over-expression of the recombinant protein when uninduced and induced samples are compared (Figure 3.4.2, leftmost black arrow). The protein also does not seem to form inclusion bodies during expression as evident by the lack of a major protein band at around 38 kDa in the insoluble fraction. The major band seen in the insoluble fraction is lower molecular weight than what is expected for $TBL1X_{176-527}$ construct. It is also present in pulldowns done on other TBL1X constructs (Figure 3.3.1) and is therefore likely to be a native *E. coli* protein. There was a band in the elution fractions (1A7-1B2) that was the expected molecular weight for the $TBL1X_{176-527}$ construct (rightmost black arrow).

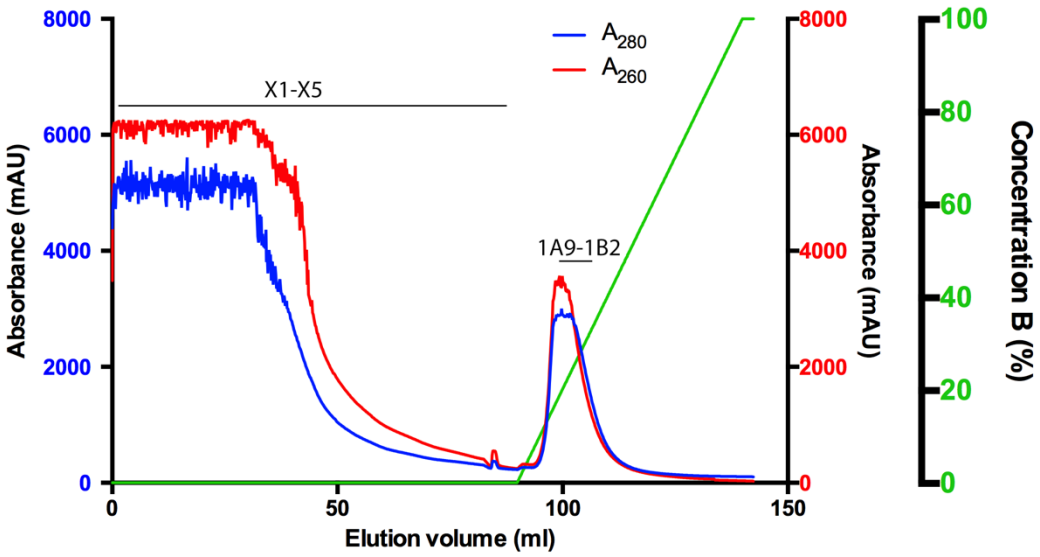

Figure 3.4.1. First immobilised metal affinity chromatography step chromatogram. The X1-X5 fractions correspond to flow-through of lysate.

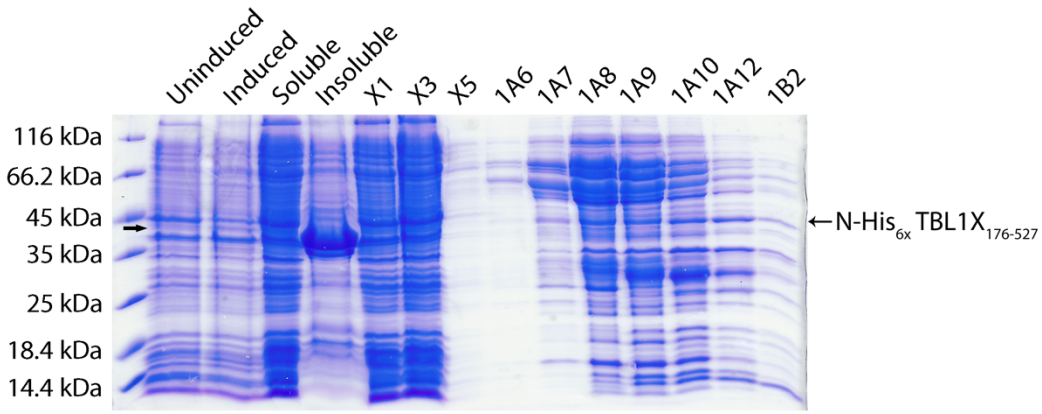

Figure 3.4.2. SDS-PAGE gel of the immobilised metal affinity step.

Fractions containing $TBL1X_{176-527}$ (1A9 to 1B2, Figure 3.4.2) were pooled and dialysed overnight into anion-exchange chromatography buffer. Fractions 1A7 and 1A8 were not pooled because of the ratio of contaminants to $TBL1X_{176-527}$, even though this resulted in the loss of some recombinant protein. 3C protease was added to the pooled fractions to remove the hexahistidine tag. After overnight buffer exchange the sample in the dialysis bag showed considerable precipitation, however upon analysis it was shown that $TBL1X_{176-527}$ was still present in the soluble fraction as indicated by altered migration of the band corresponding to $TBL1X_{176-527}$. This suggested that the hexahistidine tag had been cleaved off (Figure 3.4.3). The sample was filtered using a 0.22 µm filter and loaded onto an anion-exchange column. The flow-through was collected for analysis. Bound protein was eluted with a linear gradient

of NaCl (0-1M). Most of $TBL1X_{176-527}$ was in the flow-through and only a small amount bound to the column (Figure 3.4.3 and Figure 3.4.4). There was also some $TBL1X_{176-527}$ present in the elution fractions (Figure 3.4.3, rightmost black arrow), although that did not constitute a majority of the protein.

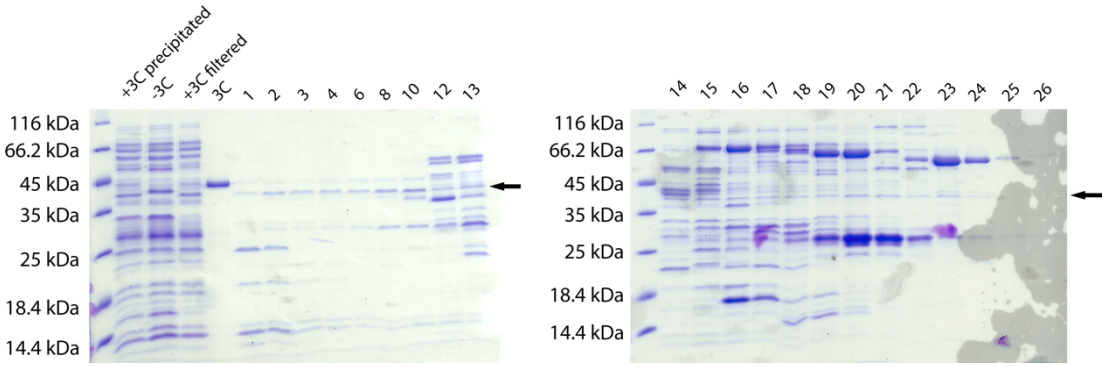

Figure 3.4.3. SDS-PAGE gel of anion-exchange step. Most of $TBL1X_{176-527}$ was in the flow-through (fractions 1-3, leftmost black arrow) while a lot of the contaminating proteins bound to the column. +3C precipitated corresponds to the overnight dialysed sample, +3C filtered is the filtered overnight sample, and -3C is a sample without any added 3C protease. 3C lane is 3C protease.

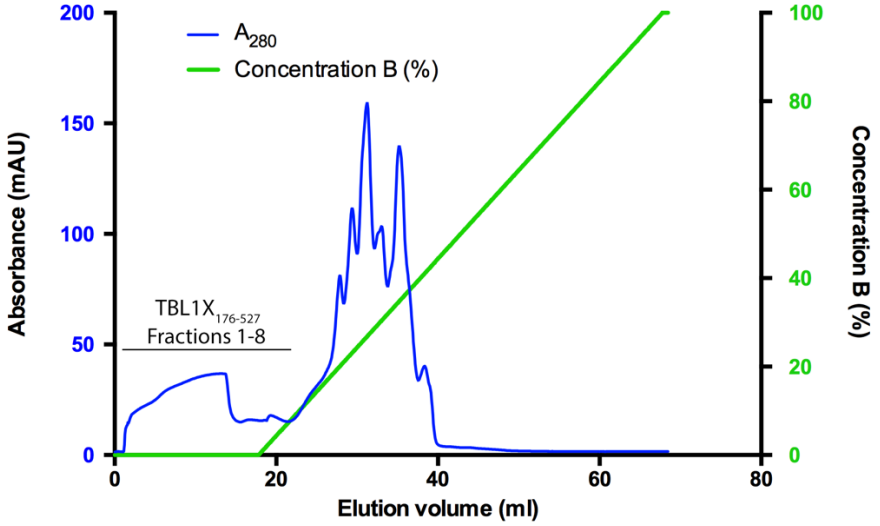

Figure 3.4.4. Anion-exchange elution profile of $TBL1X_{176-527}$. The majority of $TBL1X_{176-527}$ did not bind to the column and is present in the flow through (fractions 1-3), although some of the protein did bind and eluted without any salt (fractions 4-8). The protein also eluted all across the linear gradient, so it is unlikely that the interaction with the resin was specific.

While the anion-exchange step helped to clean out many contaminants, the sample was still relatively heterogeneous. In order to purify $TBL1X_{176-527}$ further, the flow through fractions were pooled and reloaded onto an IMAC column. IMAC columns are known for their binding

of untagged proteins due to the presence of histidine runs in native proteins. Since the $TBL1X_{176-527}$ construct does not contain polyhistidine runs in its sequence, and the hexahistidine tag had been cleaved off, it was theorised that the contaminating proteins would bind to the column while the $TBL1X_{176-527}$ would be in the flowthrough. After the second IMAC step there was detectable 280 nm absorbance both in the flowthrough as well as in the elution fractions. SDS-PAGE gel analysis showed that the $TBL1X_{176-527}$ construct did not bind to the column (A3-B9) while most contaminants did (E3-G9). One contaminant co-eluted with $TBL1X_{176-527}$. It was believed to be the 3C protease although this was not confirmed by further analysis (Figure 3.4.5 and Figure 3.4.6).

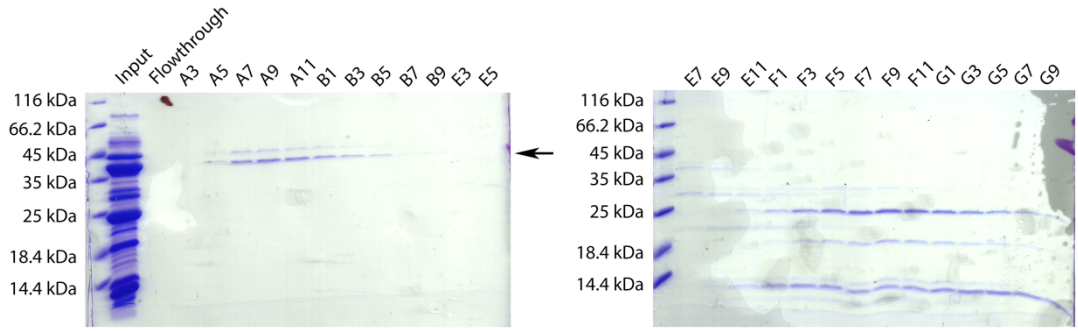

Figure 3.4.5. SDS-PAGE analysis of the second step of IMAC. $TBL1X_{176-527}$ and a contaminating band, believed to be the GST-3C protease, were present in the flow through (A3-B9, black arrow), while the contaminating proteins bound to the column (E3-G9). The flow through lane corresponds to the flow through from the 10 kDa MWCO concentrating column.

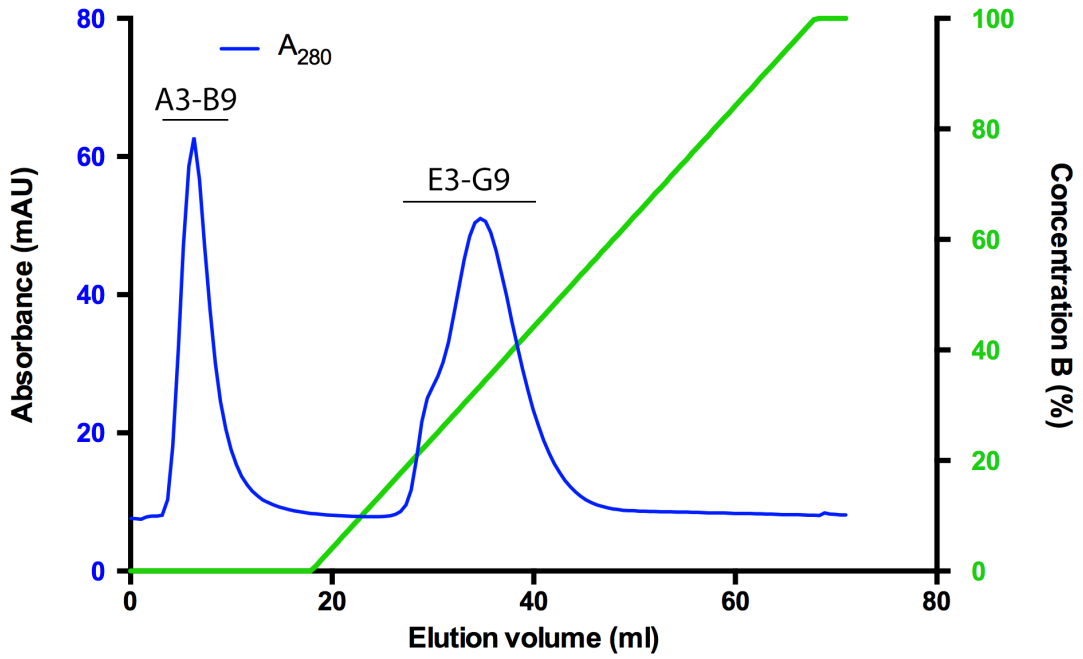

Figure 3.4.6. Second IMAC profile.

The appropriate fractions were again pooled and separated by size-exclusion chromatography. This was performed to analyse the heterogeneity of the sample, as well as to separate the protein from contaminants still present in the sample. The resulting chromatogram combined with SDS-PAGE gel analysis showed that the recombinant protein was a single species that eluted at around 16 ml (Figure 3.4.7 and Figure 3.4.8). Analysis of MALDI-TOF-TOF data performed on the bands in Figure 3.4.8 resulted in 18 peptide matches to mouse TBL1X with a Mascot score of 122. The score is given as -10*log(P), which is the probability of the match being a random event. Final yield, as determined by measuring the absorbance at 280 nm, was approximately 45 µg from 20 grams of wet cell pellet.

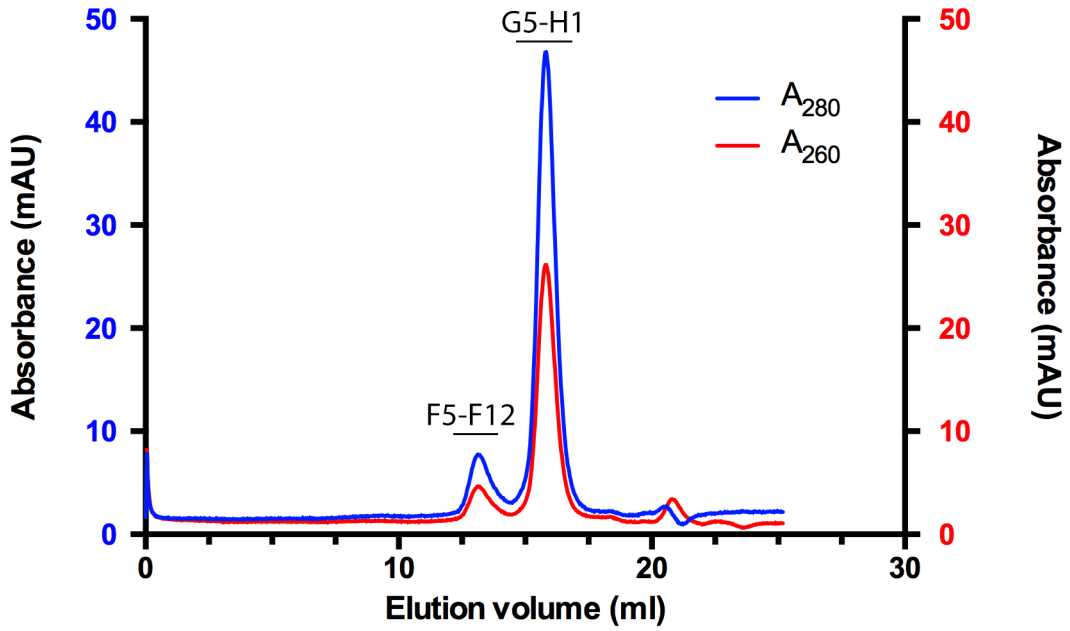

Figure 3.4.7. Size-exclusion chromatography profile for TBL1X$_{176-527}$ purification. TBL1X$_{176-527}$ eluted at around 15 ml (G5-H1). The contaminating protein (F5-F12) was successfully separated from TBL1X$_{176-527}$.

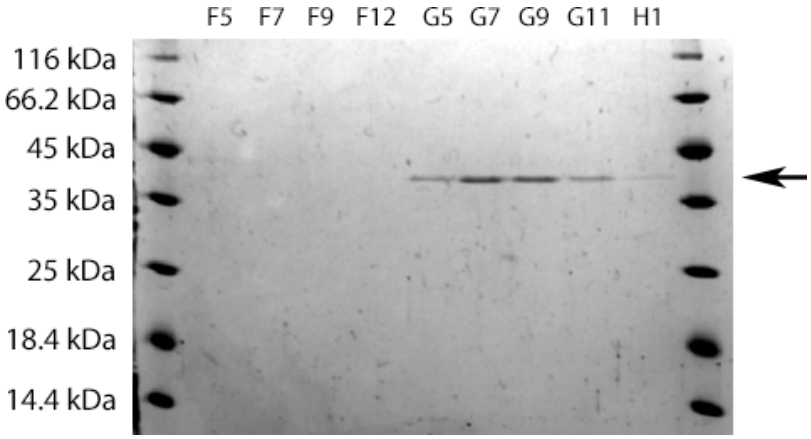

Figure 3.4.8. SDS-PAGE analysis of size-exclusion chromatography step. TBL1X$_{176-527}$ (fractions G5-H1, black arrow) was separated from the contaminant (fractions F5-F7, band too faint to be seen on a scanned gel), resulting in a clean monomeric protein.

3.5 Expression optimisation of TBL1X$_{176-527}$ in *E. coli*

The TBL1X$_{176-527}$ construct yielded soluble monomeric protein, albeit at very low levels that were insufficient for crystallisation trials or biophysical assays. Therefore, attempts were made to increase the expression levels of this construct. First, two different types of media

with two different concentrations of IPTG at different temperatures were tested. Terrific Broth sometimes shows better protein expression levels as it contains both K_2HPO_4 and KH_2PO_4 as a buffering agent, which reduces media acidification that is detrimental to *E. coli* growth. The tests were done first in BL21(*DE3*) cells with 300 µM IPTG (Table 3.5.1) and grown overnight. Pulldowns were done on overnight samples, although for each condition hourly time points up to six hours were taken and soluble fractions analysed as generally protein overexpression is visible even without first enriching for the protein of interest.

Table 3.5.1. Expression tests of TBL1X$_{176-527}$ done in BL21(*DE3*) cells using 100 and 300 µM IPTG for induction. 'No' means no protein was seen in pulldowns. 'Yes' means band corresponding to the molecular weight of TBL1$_{176-527}$ was seen in the pulldown sample, however the level was not higher than what was seen at 18˚C. 'N/A' means the expression was not tested.

100 µM IPTG	**20˚C**	**25˚C**	**30˚C**	**37˚C**
2XTY	Yes	N/A	Yes	No
Terrific Broth	Yes	N/A	Yes	No
300 µM IPTG	**20˚C**	**25˚C**	**30˚C**	**37˚C**
2XTY	Yes	Yes	Yes	No
Terrific Broth	Yes	Yes	Yes	No

When the cells were grown at 37˚C there was no TBL1X$_{176-527}$ present in the pulldowns. For all other temperatures there was a band present that was the correct molecular weight, however the levels of expression did not exceed those seen in the initial samples grown at 18˚C. A representative gel of expression test done at 25˚C has been shown in Figure 3.5.1.

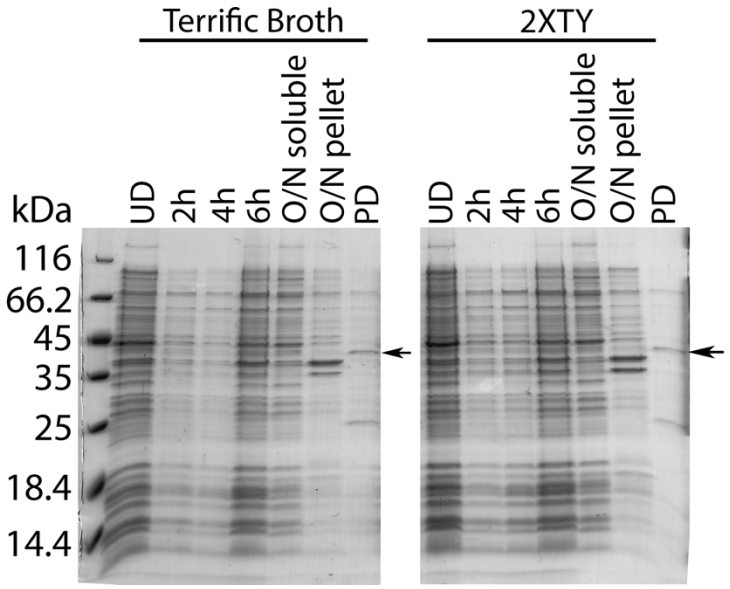

Figure 3.5.1. Expression test of TBL1$_{176-527}$ at 25˚C using 300 µM IPTG with two different types of medium. Bands corresponding to the protein of interest are marked with black arrows. The gel has been digitally cut in half to add the arrow but has not been altered in any other way.

The expression at 20˚C, 30˚C, and 37˚C were then repeated using 100 µM IPTG in hopes that lower IPTG concentration and therefore slower induction would help to make more soluble protein. Expression at 25˚C was not tested because there was no difference in recombinant protein levels between 20˚C and 30˚C using 300 µM IPTG. At 37˚C there was not any protein present in the pulldown fractions. There was a band corresponding to the right molecular weight again present in both 20˚C and 30˚C samples, however, similarly, their levels of expression were not greater than those at 18˚C. For both 100 µM and 300 µM IPTG tests the 6-hour time point did not reveal any increase in the recombinant protein levels in either soluble or insoluble fraction, suggesting that there were problems with translation rather than degradation of the protein.

Another type of medium tested was autoinducible media (Overnight Express™ Instant TB, Novagen) as it can produce higher cell density with better protein yields. While the OD_{600} values following an overnight growth were about twice as large as for normal medium (OD_{600} of 20 for autoinducible versus 10 for 2XTY), pulldowns showed no expression of TBL1X$_{176-527}$. A codon-optimised version of the TBL1X$_{176-527}$ construct was synthesised to see whether the low levels of expression were due to rare codon usage in the mammalian coding sequence. According to codon analysis (Rare Codon Calculator), 24 out of the 372 (~6.5%) codons of the mouse coding sequence were annotated as being rarely used in *E. coli.* However, when

tested, the codon-optimised construct had abolished the expression of the protein completely.

After the failure of the codon-optimised gene to improve expression of the protein, combined with the lack of success of varying different parameters in the expression tests, it was decided not to pursue further expression optimisation in *E. coli*.

3.6 Expression tests in Spodoptera frugiperda

While the expression of TBL1X in *E. coli* worked in principle, the yields were low, making it difficult to produce protein in sufficient quantities for biochemical assays and attempts to increase the yields were unsuccessful. Attention was turned towards insect cells and baculovirus expression system as the human TBL1XR1 WD40 domain that had been crystallised (PDB ID: 4LG9) was expressed in *Spodoptera frugiperda* Sf9 cell line.

Six constructs were cloned into insect cell expression vectors; three in TBL1X and three in TBL1XR1 that started from equivalent amino acid positions (Table 3.6.1). $TBL1X_{176-527}$ and $TBL1XR1_{163-514}$, which were constructs starting at equivalent amino acids in these two proteins, were the minimal constructs that interacted with MeCP2. The $TBL1X_{148-527}$ and corresponding $TBL1XR1_{134-514}$ were based on the crystallised human TBL1XR1 WD40 domain construct (PDB ID: 4LG9). Full-length TBL1X and TBL1XR1 were chosen for the purposes of reconstituting NCoR/SMRT core co-repressor complex at a later date.

Initially only the $TBL1X_{148-527}$ and $TBL1X_{176-527}$ construct, as well as the full-length TBL1XR1 and $TBL1XR1_{163-514}$ construct were tested for expression. The plasmids of these four constructs were transformed into *E. coli* cells containing the baculovirus genome in an artificial chromosome (bacmid). Following the integration of the plasmid into the bacmid, the bacmid was isolated and used for the generation of the baculovirus stocks. The by-product of virus generation in Sf9 cells was a pellet of infected, protein-expressing cells used for pulldown assays.

Table 3.6.1. All insect cell constructs that were created for the purposes of protein expression. The equivalent construct of TBL1X$_{148-527}$ was TBL1XR1$_{134-514}$, and the equivalent construct of TBL1X$_{176-527}$ was TBL1XR1$_{163-514}$.

Protein	Construct	Molecular weight (kDa)	Tag	Organism	Vector
TBL1X	1-527	59.0	N-His$_{6x}$	*M. musculus*	pFL
TBL1X	148-527	43.7	N-His$_{6x}$	*M. musculus*	pFL
TBL1X	176-527	40.7	N-His$_{6x}$	*M. musculus*	pFL
TBL1XR1	1-514	58.0	N-His$_{6x}$	*M. musculus*	pFL
TBL1XR1	134-514	43.9	N-His$_{6x}$	*M. musculus*	pFL
TBL1XR1	163-514	40.7	N-His$_{6x}$	*M. musculus*	pFL

Another cell line that is often used in baculovirus expression system is the High Five™ (Hi5) from *Trichopulsia ni*. For some proteins the expression in Hi5 cells can be many times higher than in Sf9 cells, making them attractive for protein expression. A common problem is the tendency of these cells to clump together which often results in cell death. This can be overcome by adding dextran to the medium, although this poses another challenge as dextran sometimes interferes with the infection of the cells by the baculovirus. Nevertheless, infected cells from both cell lines were used for small-scale pulldown experiments.

As can be seen from Figure 3.6.1, for both Sf9 and Hi5 cell lines the TBL1X$_{148-527}$ and TBL1X$_{176-527}$ constructs, as well as the full-length TBL1XR1 and TBL1XR1$_{163-514}$ constructs expressed in both cell lines to different levels, with expression being generally higher in Sf9 cells. They also revealed that the expression of these proteins was much higher in insect cells than it was in *E. coli.* These experiments were repeated multiple times and each time the expression of protein in Hi5 cells was highly variable while it was consistent in Sf9 cells. To confirm that the proteins seen in the pulldowns were indeed various constructs of TBL1X and TBL1XR1, Western blots were carried out using an antibody against the C-terminus of TBL1XR1, which is able to recognise both TBL1X and TBL1XR1. As can be seen from Figure 3.6.2, the overexpressed protein seen on the SDS-PAGE gels is recognised by the anti-TBL1XR1 antibody.

While time was spent trying to understand the nature of the variation in Hi5 cells by adjusting the variables in the infection protocol, the variable expression problem was never resolved. Due to this, all subsequent protein expression tests, as well as large-scale protein expression, were carried out in Sf9 cells.

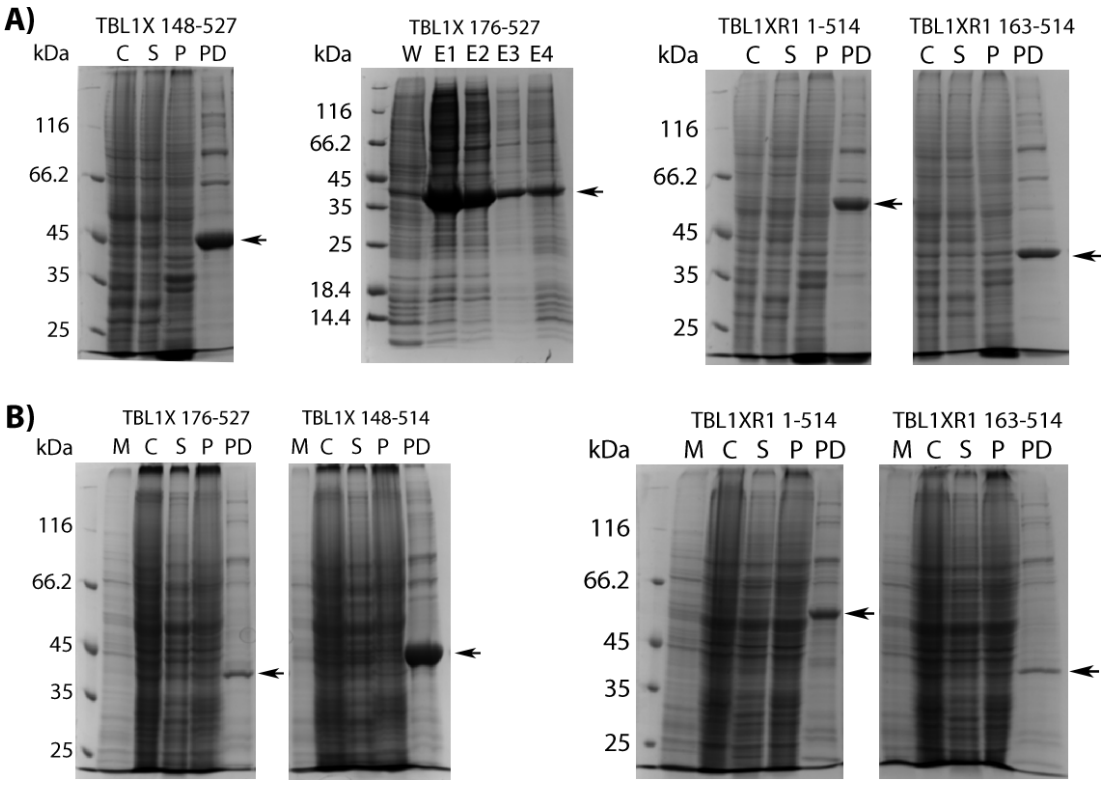

Figure 3.6.1. SDS-PAGE gels of expression tests of cloned TBL1X and TBL1XR1 constructs in Sf9 (A) and Hi5 cells (B). Bands corresponding to the protein of interest have been marked with a black arrow. Lanes are marked as follows: C - Crude lysate, S - Soluble fraction, P - Pellet, PD - Pulldown, M - Medium, W - wash, E1-E4 - Elution 1-4. Due to a loading error, there was no pulldown for $TBL1X_{176-527}$, but purification trials at a later date showed that the protein expressed to high levels ($TBL1X_{176-527}$ Sf9 gel). While all the constructs expressed in both Sf9 and Hi5 cells, there was considerable variation between same constructs in Sf9 and Hi5 cells. Gels have been digitally cut in half to add the arrows but have not been altered in any other way.

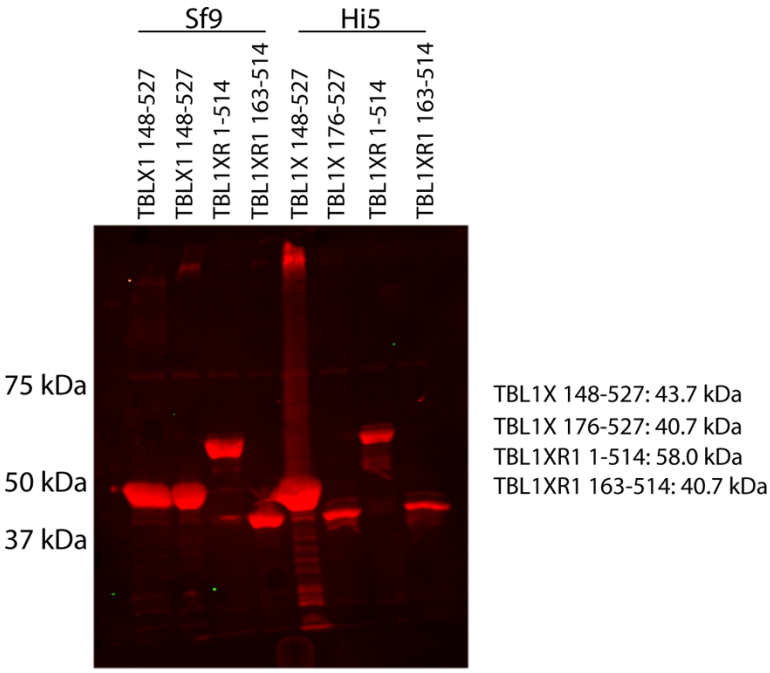

Figure 3.6.2. Western blots of pulldown fractions seen in Figure 3.6.1 using an antibody against TBL1XR1 C-terminus which is also able to recognize TBL1X. TBL1X$_{148-527}$ appears twice, again due to loading error from previous experiment. Expected molecular weights of the proteins are on the right. All the bands seen correspond to the expected molecular weights for the TBL1X and TBL1XR1 constructs.

3.7 Expression optimisation trials in Sf9 cells

While the protein expressed well in Sf9 cells as judged from the SDS-PAGE gels of the pulldowns, it was decided to try to increase the protein expression levels. Two factors that can affect the protein yield are cell density at infection and amount of virus. To assess the importance of these two variables, a range of expression tests were set up with cell counts ranging from $2*10^6$ cells/ml to $4*10^6$ cells/ml and virus:medium ratio ranging from 1:10 to 1:50. The cell viability was measured using a cell counter every day for three days, and the fluorescence of the cells was monitored using a fluorescence microscope. Regardless of the initial cell count or amount of virus, the optimal time to harvest cells was after three days as, after that, the viability started to drop, indicating cell death and lysis. The optimal virus to medium ratio was between 1:20 and 1:50 depending on the virus stock used, although generally the V2 stocks were high enough in titre to be used at 1:50 dilution. Using more virus than 1:20 (e.g. 1:10) resulted in very quick growth arrest and cell death due to high rate of infection. Because the Sf9 cells continued growing for a while after being infected, it was

hoped that using a higher initial count of cells would yield more cell mass, and thus more protein, by the end of the infection. However, using $4*10^6$ cells/ml just resulted in more cells being uninfected and/or dead despite varying the amount of virus used. This was hypothesized to occur due to either nutrient depletion or crowding of the cells, although this was not examined further.

Although each V2 virus stock generated was different in the virus titre, a standard protocol for large-scale protein expression was used. This meant using $2*10^6$ cells/ml and infecting them with a virus:medium ratio of 1:50. If a particular V2 stock happened to show poor infectivity with these parameters, the virus:medium ratio was increased to 1:40 or even 1:20. If there were still problems with infection, the V2 stock was remade from V1. Finally, if the new V2 stock still showed problems with infectivity, the entire virus stock for that particular construct was discarded and fresh V0 was generated.

3.8 Establishing purification protocol of TBL1X/TBL1XR1

Initially the purification protocol that had been developed for bacterial TBL1X purification was taken as a starting point for $TBL1X_{148-527}$. Following cell lysis and clearing of the lysate by centrifugation, the supernatant was loaded onto an IMAC column. The column was washed to remove any unbound protein, and bound protein eluted with a gradient of 0-250 mM imidazole. As can be seen from the chromatogram and the corresponding SDS-PAGE gel (Figure 3.8.1, Figure 3.8.2), the protein bound to the column and it was possible to separate the protein of interest from most other contaminants (fractions 18-22).

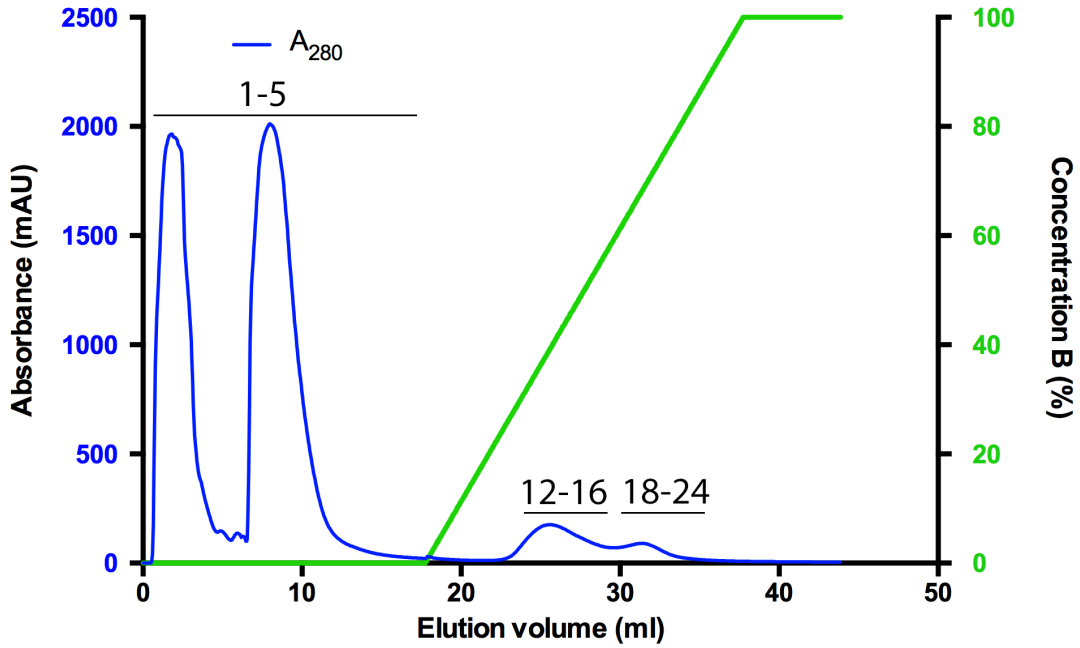

Figure 3.8.1. IMAC chromatography step profile of TBL1X$_{148-527}$.

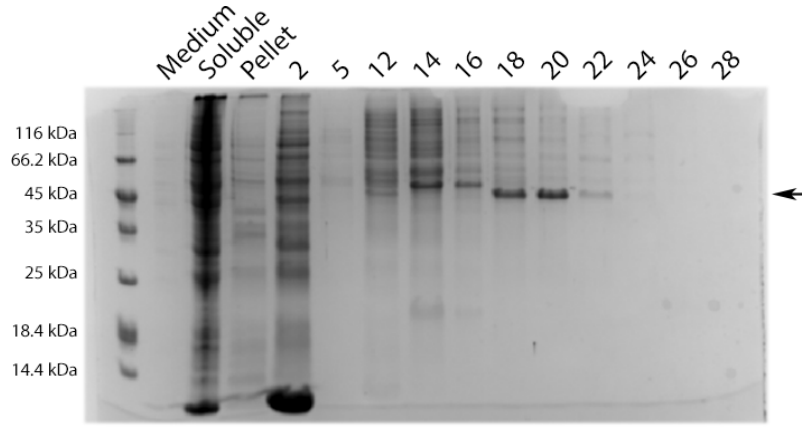

Figure 3.8.2. SDS-PAGE gel of the first IMAC purification step of TBL1X$_{148-527}$. TBL1X$_{148-527}$ (marked with black arrow) bound to the column and eluted separately (fractions 18-22) from most contaminants (fractions 12-16).

There were some contaminants present so in order to separate TBL1X from these, a cation exchange column was used. Previously an anion exchange column had been used in the purification attempts with the *E. coli* protein as, according to ProtParam, the predicted pI of the protein was 5.96 and would thus would carry an overall negative charge in the buffer used (pH 7.5). However, that step resulted in only a partial recovery of protein, so it was investigated whether it was possible to clean up the protein with a cation-exchange column instead by allowing the contaminants to bind and collecting the protein of interest from the flow-through. The fractions containing TBL1X$_{148-527}$ (18-22) were pooled together and

concentrated. The protein was resuspended in buffer containing low salt and then applied to the column. However no protein was detected in the flow-through and no protein eluted off the column using a 0.1-1M NaCl gradient (Figure 3.8.3, Figure 3.8.4). Although no increase in the pressure over the column was seen, this still implied that the protein had aggregated on the column. Indeed, when the column was subsequently washed with 0.5 M NaOH, a large peak was observed eluting from the column that most likely corresponded to the applied $TBL1X_{148-527}$. Although this peak was not analysed on an SDS-PAGE gel, the column was cleaned thoroughly prior to use and the only protein that was present on the column was the applied sample.

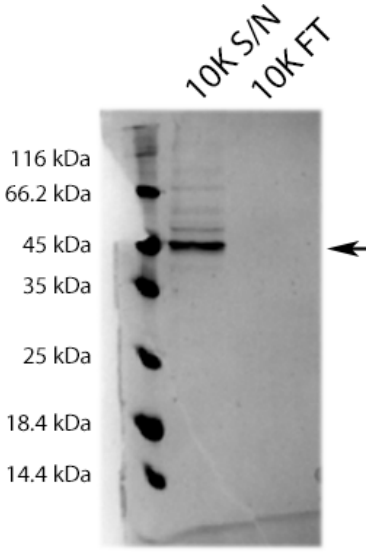

Figure 3.8.3. SDS-PAGE gel of $TBL1X_{148-527}$ after concentrating with a 10K molecular weight cut-off (MWCO) concentrator. As is evident, all of the protein of interest (marked with a black arrow) stayed in the top part of the concentrator, while no detectable amount of protein was in the flow-through (10K F/T). The supernatant (10K S/N) fraction was applied to the cation exchange column in the next step of purification.

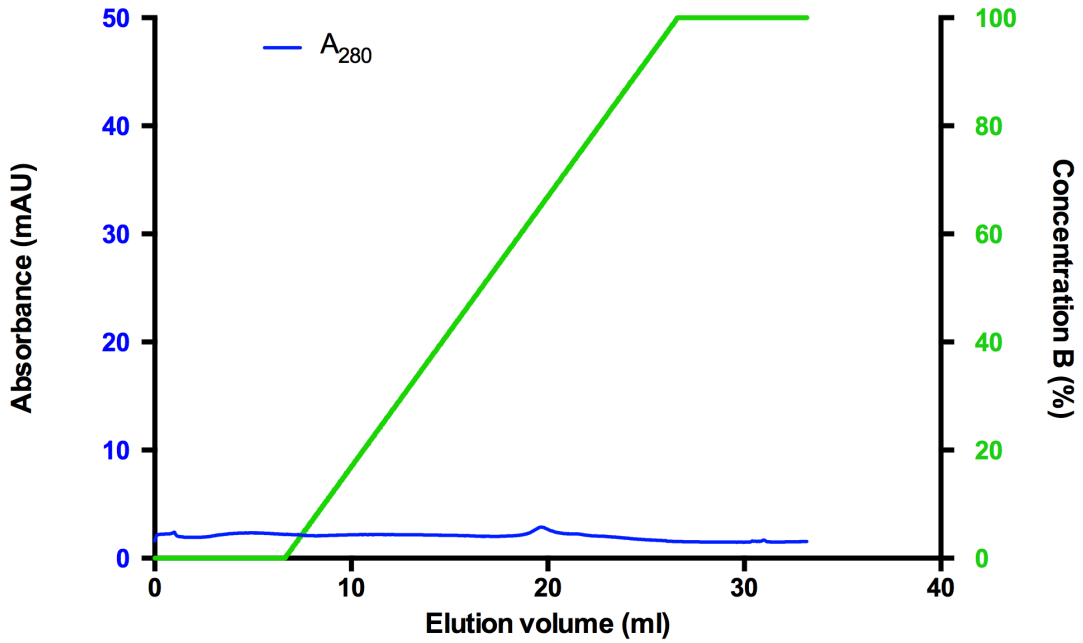

Figure 3.8.4. Cation-exchange chromatography profile for TBL1X$_{148-527}$ construct.

It was not immediately clear what caused the protein to aggregate on the column. However as this resulted in total loss of protein it was decided to remove the ion-exchange step and see whether size-exclusion chromatography would be an appropriate next purification step. An issue with using insect cells was the difficulty of cleaning the supernatant from cell membranes and other larger species through centrifugation at high speeds. This debris can cause affinity columns to block after a certain amount of cleared lysate had been loaded. Attempts to clear the supernatant using ultracentrifugation were mostly unsuccessful. While some cellular debris was pelleted during ultracentrifugation, the applied sample still clogged the column eventually. Attempts to dilute the lysate to make it less viscous did not produce any positive results. Therefore, in order to make the protocol easier to execute, the initial IMAC step was switched from using a column to using Ni^{2+}NTA resin in batch. This had the additional benefit of increasing the recovery of the protein, most likely due to increased contact surface and time (Figure 3.8.5).

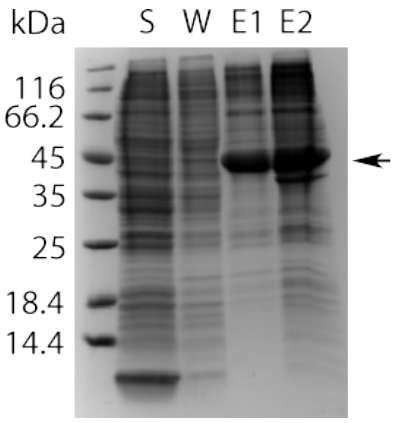

Figure 3.8.5. SDS-PAGE gel of IMAC step in the purification of TBL1X$_{148-527}$. This step was done in batch which resulted in a considerable increase in the yield of protein of interest (black arrow). S – soluble fraction. W – wash. E1 and E2 – elutions 1 and 2.

The first and second elutions were dialysed overnight into gel-filtration buffer. This step again resulted in considerable precipitation developing in the dialysis bag. However, as before, TBL1X$_{148-527}$ remained in solution. The overnight sample was pooled, filtered, and concentrated after which it was injected over a gel-filtration column.

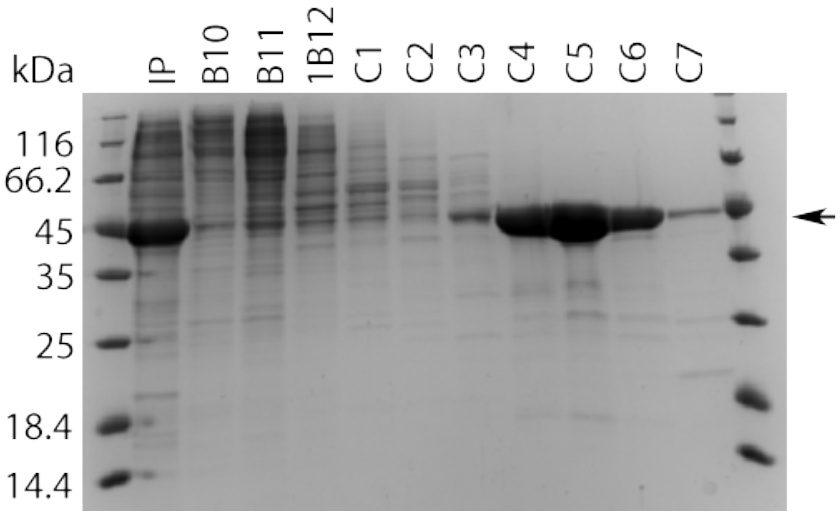

Figure 3.8.6. SDS-PAGE of size-exclusion chromatography step. TBL1X$_{148-527}$ eluted near its expected size (black arrow). IP – input.

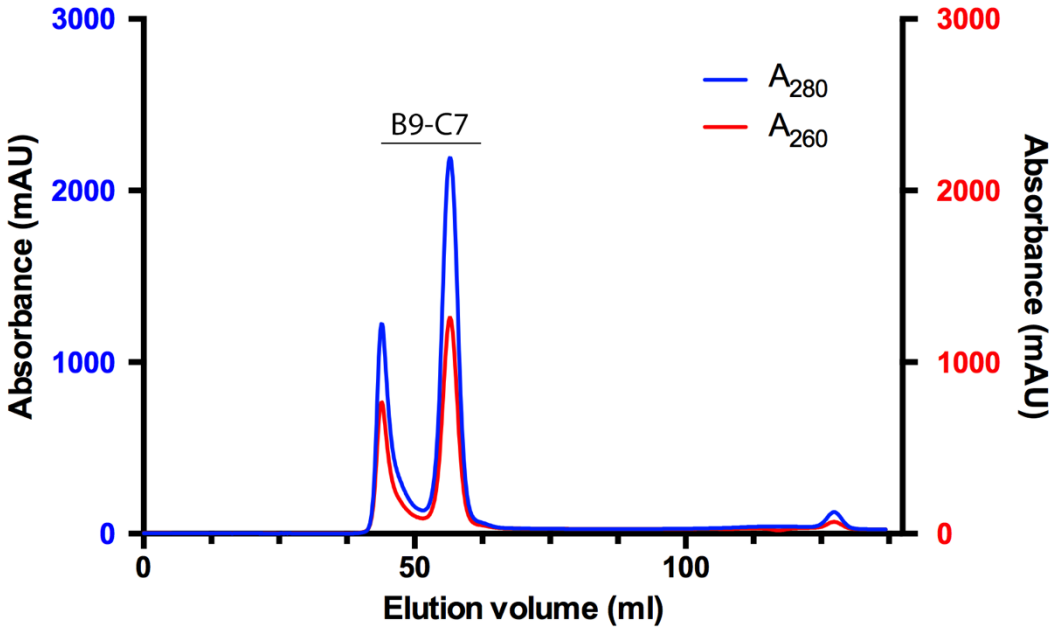

Figure 3.8.7. Size-exclusion chromatography profile of TBL1X$_{148-527}$ purification.

As can be seen from Figure 3.8.6 and Figure 3.4.7, there were two peaks present. The earlier one around 45 ml corresponded to the void volume and consists of larger proteins and aggregates. The second monodisperse peak around 55 ml contained mostly TBL1X$_{148-527}$ (fractions C4-C6). As can be seen from the A_{260}/A_{280} ratio, there was virtually no contaminating DNA/RNA. This was confirmed by measuring the ratio using a NanoDrop spectrophotometer. The purity of the protein was more than 95% as estimated from the gel. Since there seemed to be almost no protein of interest in the void fractions, and the protein that was recovered was more than pure enough for crystallisation studies, it was decided to implement this protocol as standard.

The only change to this established protocol was that the collected fraction size was changed from 2 ml to 1 ml, resulting in better separation of some of the earlier contaminants (fractions E3-E6) while, as expected, not affecting the protocol in any other way (Figure 3.8.8).

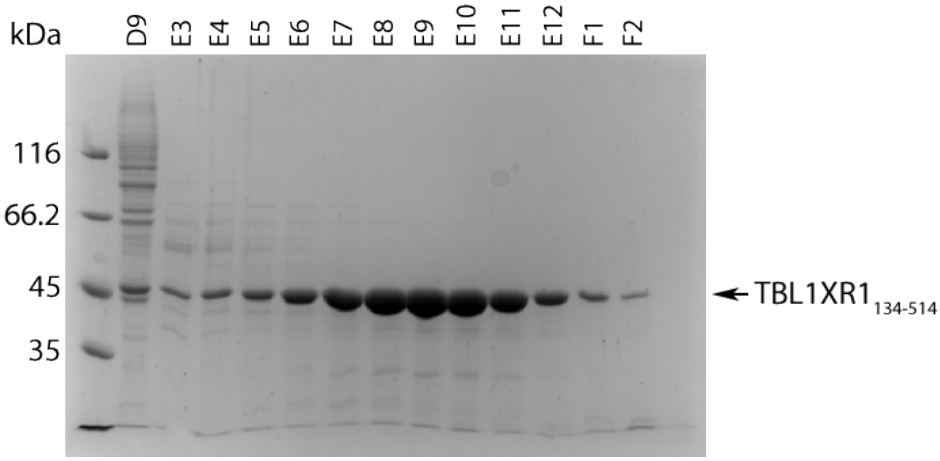

Figure 3.8.8. Representative size-exclusion chromatography SDS-PAGE gel for TBL1XR1$_{134-514}$.

The final yields were normally around 4-5 mg of protein per 20 grams of wet pellet. This protocol was tested using TBL1XR1$_{134-514}$ and TBL1XR1$_{163-514}$ as well as on TBL1X$_{176-527}$ construct and all yielded similar amounts of protein with comparable purity.

3.9 Characterising TBL1X/TBL1XR1 WD40 domain stability through thermal denaturation assay

To assess the stability of the purified protein and determine the effect of additives and different buffers, a number of thermal denaturation assays were performed. This assay tracks the unfolding of the protein in response to increasing temperatures, and the subsequent exposure of hydrophobic core by using a fluorescent dye (SYPRO Orange) that binds to these exposed areas and gives off fluorescence. The changes in protein stability can be tracked by measuring the melting temperature in response to different buffer conditions (Figure 3.9.1).

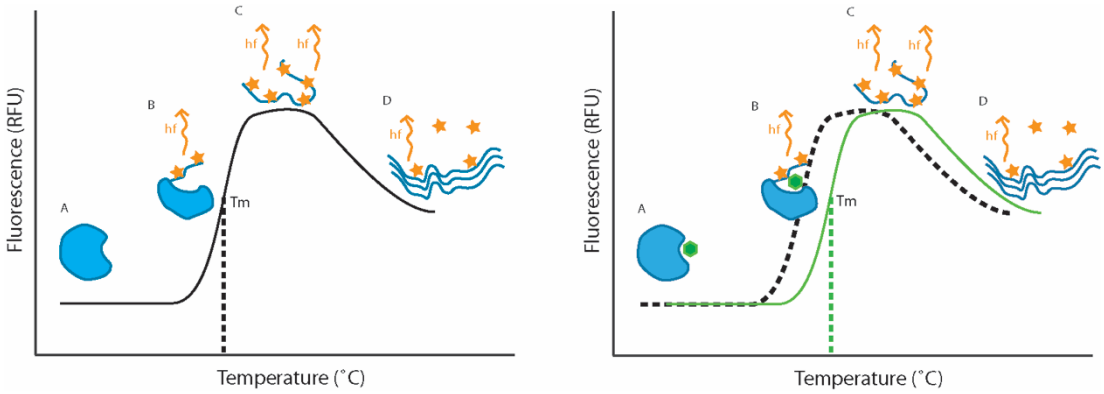

Figure 3.9.1. Hypothetical thermal denaturation assay melting curve of a protein. Left) Protein (blue) is fully folded at lower temperatures (A) and the fluorescence signal is quenched. As temperature increases, parts of the protein become unfolded and SYPRO Orange (orange pentagon) is able to

bind to hydrophobic areas (B). This allows SYPRO Orange to fluoresce. At a certain temperature all of the protein has been unfolded and maximum fluorescence has been reached (C). As the temperature increases, aggregation of protein molecules and inaccessibility of hydrophobic areas decreases the fluorescence signal (D). The melting temperature is defined as the point where the concentrations of unfolded and folded species are equal, denoted with a Tm on the graph. Right) As on the left, except addition of a stabilising ligand (green hexagon) increases the melting temperature of the protein, which is observed as a positive shift of the curve.

The effect of pH on the stability of $TBL1X_{148-527}$ was screened with buffer pH ranging from 4.5 to 10.0 (Figure 3.9.2). Four conditions (Sodium citrate/citric acid pH 4.5, 5.0, 5.2 and 5.5, data not shown) caused either aggregation or denaturing of the protein as was evident by high initial fluorescence readings, indicating exposed hydrophobic areas. This made analysis of the data to get a melting temperature impossible. This is in line with observed decrease in stability of the protein between pH 6.0 and pH 5.8 as indicated by the drop in the melting temperature from 59.7˚C to 46.7˚C. It is possible that this drop is due to the pH approaching the predicted pI of the protein. On the other hand, higher pH values, even up to 10.0, did not change the stability of the protein to a large extent. There was a slight decrease in the melting temperature after pH 9.0, but this effect was nowhere near as great as what was seen between the transition from pH 6.0 to 5.8.

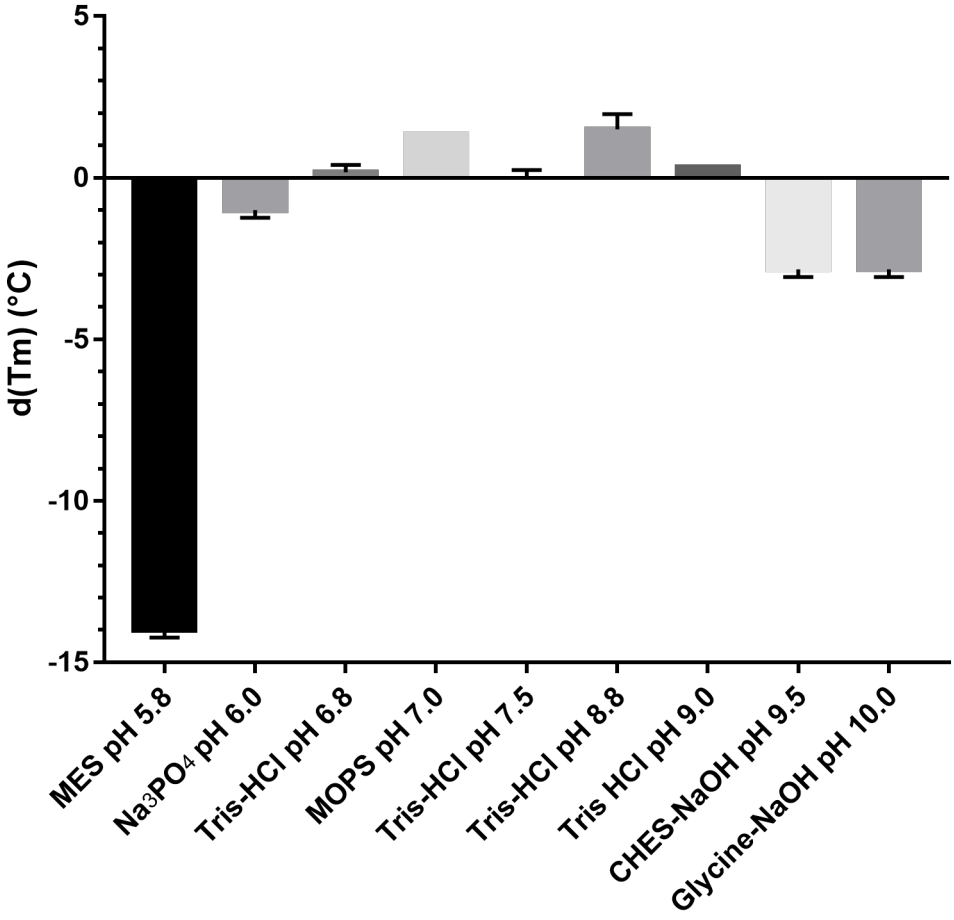

Figure 3.9.2. Stability of TBL1X$_{148-527}$ in different buffers. Melting temperature of Tris-HCl pH 7.5 sample (60.7±0.24°C) was set as the reference. The error bars are one standard deviation.

Another screen assessed the effects of different salts on the stability of TBL1XR1$_{134-514}$ (Figure 3.9.3). Increasing concentrations of Na^+, NH_4^+, and SO_4^{2-} ions increased the melting temperature of TBL1XR1$_{134-514}$. However, while lowering NaCl concentration slightly destabilised the protein and increasing NaCl stabilised it, the effect was minimal. Increasing ammonium sulphate concentration from 100 mM to 400 mM changed the melting temperature of the protein from +1.67˚C to +5.67±0.4°C when compared to the reference containing sodium chloride. Mg^{2+} ions increased the stability of the protein at low concentrations but decreased it at higher. Li^{2+} and NH_4^+ ions affected stability of the protein to a similar extent, as evidenced by melting temperatures of the protein in 200 mM $(NH_4)_2SO_4$ and 200 mM $LiSO_4$. Even though the melting temperature of the protein in most conditions was already quite high (~61-63˚C), ammonium sulphate showed a large stabilisation effect

over sodium chloride. Together these two experiments provided a range of conditions that could be used for crystallisation trials.

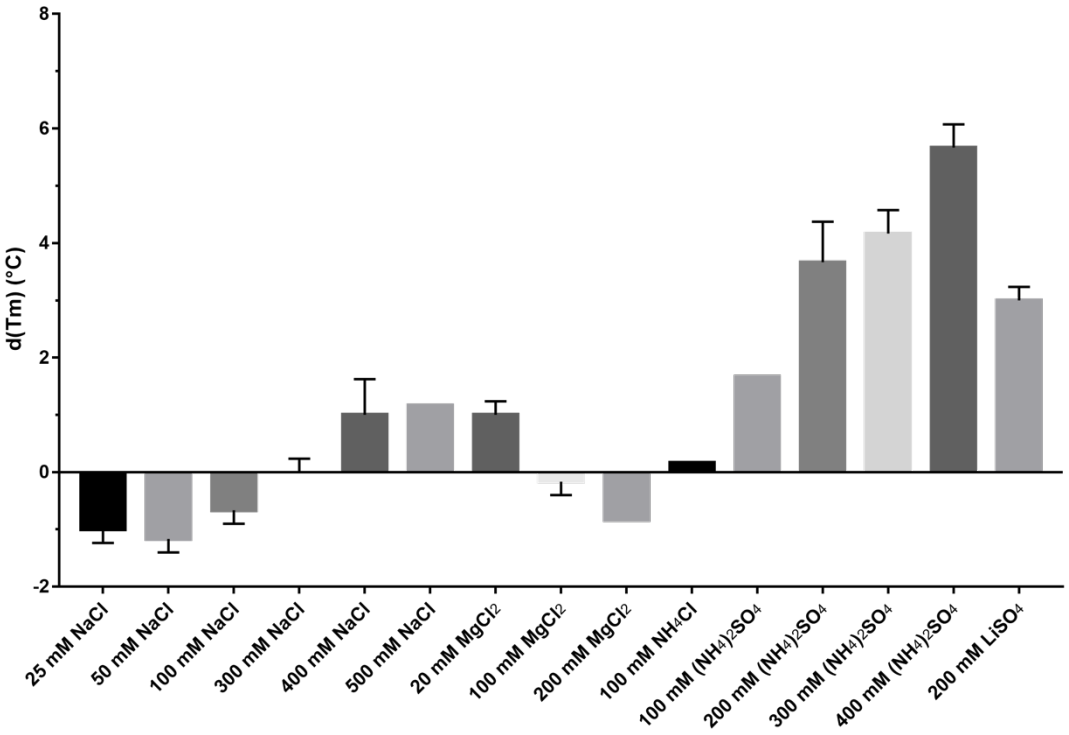

Figure 3.9.3. Stability of TBL1XR1$_{134-514}$ in different salt solutions. 300 mM NaCl sample (62.8±0.24°C) was set as the reference. When different salts were tested, NaCl was removed from the samples. The error bars are standard deviation.

3.10 Discussion

While some of the TBL1X constructs that were cloned into bacterial expression vectors expressed protein in *E. coli,* and it was possible to purify at least one of them, the final yields that were achieved were too low for any biochemical assays. Despite attempts to increase the yield through the use of different strains and temperatures, as well as varying induction time and IPTG concentration, and even trying a codon-optimised DNA, the yields did not increase. Therefore, the expression system was switched to baculovirus expression, which showed greatly increased expression levels compared to *E. coli*.

Successful purification protocols were established for TBL1X and TBL1XR1 constructs. While initially the protocol consisted of IMAC, followed by ion-exchange and then finally size-exclusion chromatography (SEC), the ion-exchange step was removed for two reasons. First, using an anion-exchange column resulted in some of the protein binding to the column and

some of it coming out in the flow-through. Furthermore, the protein then eluted throughout the salt gradient, indicating that the binding was non-specific. When cation-exchange column was used, there was no recovery of the protein either in the flow-through or in the elution fractions. Only when the column was washed with 0.5M NaOH to get rid of precipitated proteins did a peak appear. This indicated that the protein was aggregating on the column. Indeed, this was a problem at a later date in surface plasmon resonance assays, as will be discussed in chapter 5.

When the IMAC affinity step was followed by SEC, it resulted in the recovery of monomeric protein that was more than 95% pure with virtually no DNA or RNA contamination as assessed by the A_{260}/A_{280} ratio. As the SEC also served as quality control step, this protocol was set as the standard for future purifications.

Finally, the stability of the protein was assessed through thermal denaturation assays. It was shown that the protein was relatively stable between pH 6.0 and 10.0, with a melting temperature around 61-63˚C. While higher pH values were not probed, lowering the pH below 6.0 resulted in a dramatic decrease in the melting temperature of the protein, and lowering the pH to 5.5 or lower resulted in protein aggregation, at least in the buffer conditions tested. The effects of different salts at different concentrations were also probed. It was shown that while Na^+, Cl^-, NH_4^+ and SO_4^{2-} were stabilising at increasing concentrations, addition of high concentrations of Mg^{2+} had a destabilising effect on the protein. Taken together, these experiments provided valuable information for crystallisation experiments by revealing which conditions to favour and which to avoid.

4 Characterisation of TBL1X/TBL1XR1 and MeCP2 interaction

4.1 Introduction

Prior to the start of this project there was limited biochemical characterisation of the interaction between TBL1X/TBL1XR1 and MeCP2. It had been shown that a discrete region on the MeCP2 called the NCoR/SMRT Interaction Domain (NID) interacts with TBL1X WD40 domain based on pulldown experiments with cell lysates (Lyst et al., 2013). It was also shown that RTT mutations abolish this interaction, although the structural basis for this was unknown. To increase the probability of getting a co-crystallised TBL1XR1-MeCP2 complex, it was important to better understand the nature of the interaction between MeCP2 and TBL1XR1. Previous experiments carried out in the Bird lab revealed a residue in TBL1X, E364, that was critical for the interaction with MeCP2, as mutating that residue to glycine abolished the binding. Analysis of the crystal structure of human TBL1XR1 (PDB ID: 4LG9) revealed that the surface location of this residue had a strongly electronegative character (Figure 4.1.1B). As the NID was enriched in lysines and arginines, it led us to speculate that the interaction might be, at least partially, governed by electrostatics. Unpublished pulldown experiments done by Dr. Matthew Lyst also suggested that the interaction between MeCP2 and NCoR/SMRT co-repressor complex (specifically with TBL1X/TBL1XR1) was likely to only occur near physiological salt concentrations.

In this chapter I tested the hypothesis that the MeCP2-TBL1XR1 interaction is governed by electrostatic interactions. The effects of RTT mutations on the interaction between TBL1X/TBL1XR1 and MeCP2 were also investigated using *in vitro* pulldowns and thermal denaturation assays with purified recombinant proteins. Finally, the affinity of this interaction was measured using surface plasmon resonance (SPR) and fluorescence anisotropy assays (FA)

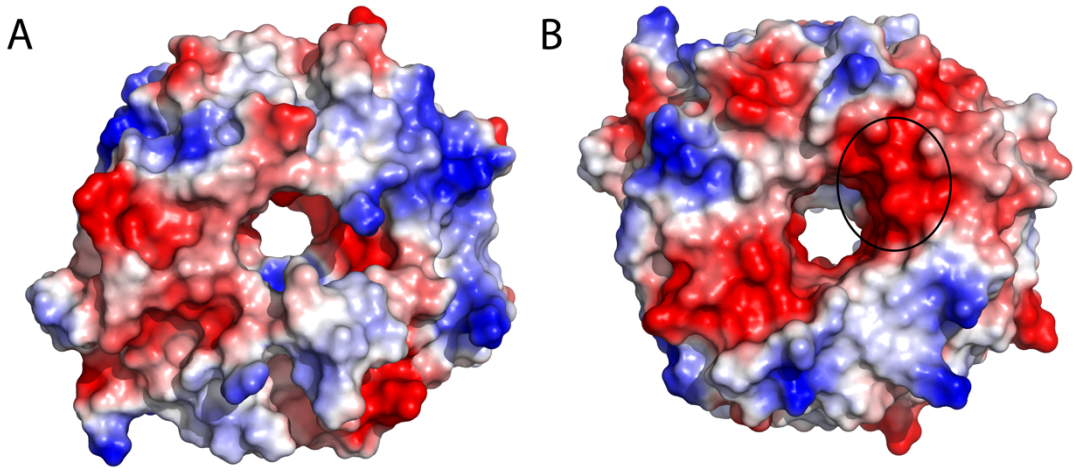

Figure 4.1.1. Electrostatic map of the human TBL1XR1 (PDB ID: 4LG9) A) bottom face and B) top face generated in PyMol. Areas with predominantly electronegative characteristic are coloured in red while electropositive areas are coloured in blue. The area containing the E364 residue that was found to be critical for mediating interaction with MeCP2 has been circled. This area had a pronounced electronegative character.

4.2 MeCP2 interacts with TBL1XR1 in a salt-dependant manner *in vitro*

The effect of increasing concentrations of NaCl and $(NH_4)_2SO_4$ on the interaction between $MeCP2_{285-309}$ and $TBL1XR1_{134-514}$ (Referred to as both TBL1XR1 and $TBL1XR1_{134-514}$ in this chapter) was probed using thermal denaturation assay. As the protein purification buffer contained 200 mM NaCl, it was necessary to establish whether MeCP2 bound to TBL1XR1 in these conditions, or whether the buffer conditions needed to be changed. The effect of $(NH_4)_2SO_4$ was also probed, as ammonium sulphate is a common precipitation agent in protein crystallisation. The effect of the salt on the protein alone, as well as on the protein and peptide interaction was analysed. If the conditions tested favoured binding of the wild type MeCP2 peptide to TBL1XR1, an increase in the melting temperature would likely be observed between protein sample and protein with peptide. However, if there was no binding, the thermal stability of the two samples should be the same.

As had been established previously, low concentrations of NaCl did not cause destabilisation of the TBL1XR1 WD40 domain. Therefore, a series of salt concentrations were tested, ranging from 25 mM to 400 mM NaCl. The thermal stability of the protein alone increased only

slightly (~3-4˚C) as the concentration of NaCl increased, in line with previous experiments (Figure 4.2.1). However, upon the addition of the $MeCP2_{285-309}$ peptide in 25 mM NaCl, the stability of TBL1XR1 increased by 16.5˚C. Increasing concentrations of salt had a diminishing effect on the thermal stability of the protein-peptide complex and at 400 mM NaCl there was virtually no difference in the thermal stability of TBL1XR1 in the absence or presence of MeCP2 peptide. This indicated, indirectly, that there was very little to no interaction between the two proteins at high salt concentrations.

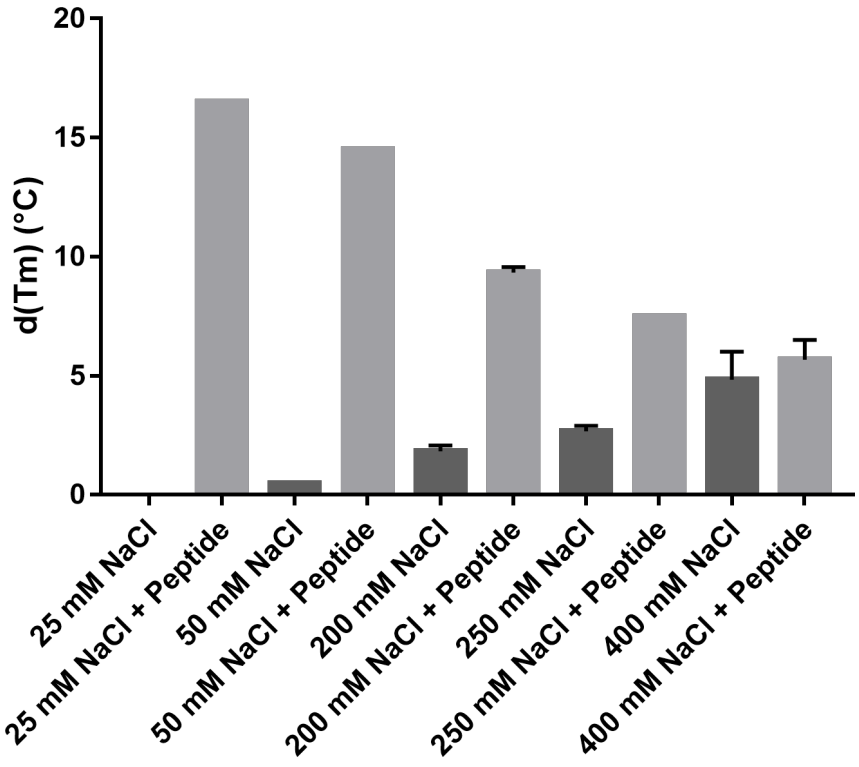

Figure 4.2.1. Binding of $TBL1XR1_{134-514}$ to $MeCP2_{285-309}$ in the presence of increasing concentration of NaCl. TBL1XR1 in 25 mM NaCl was set as the reference (61˚C). Error bars are one standard deviation.

A similar effect was seen with $(NH_4)_2SO_4$ (Figure 4.2.2): in the presence of 100 mM $(NH_4)_2SO_4$ a 3˚C increase in the thermal stability of TBL1XR1 was observed upon peptide addition, although the magnitude of this change was smaller than what was observed with similar concentrations of NaCl. However, when 200 mM $(NH_4)_2SO_4$ was added there was very little difference between the melting temperatures of the protein alone and the protein and peptide. This implied, again indirectly, that there was very little to no binding of MeCP2 peptide to TBL1XR1 in high salt. This was valuable information as it suggested that a standard

approach to sparse matrix screening was unlikely to work. This is because many commercial crystallisation screens use high concentrations of salt as a precipitant, which would abolish the interaction between the two proteins.

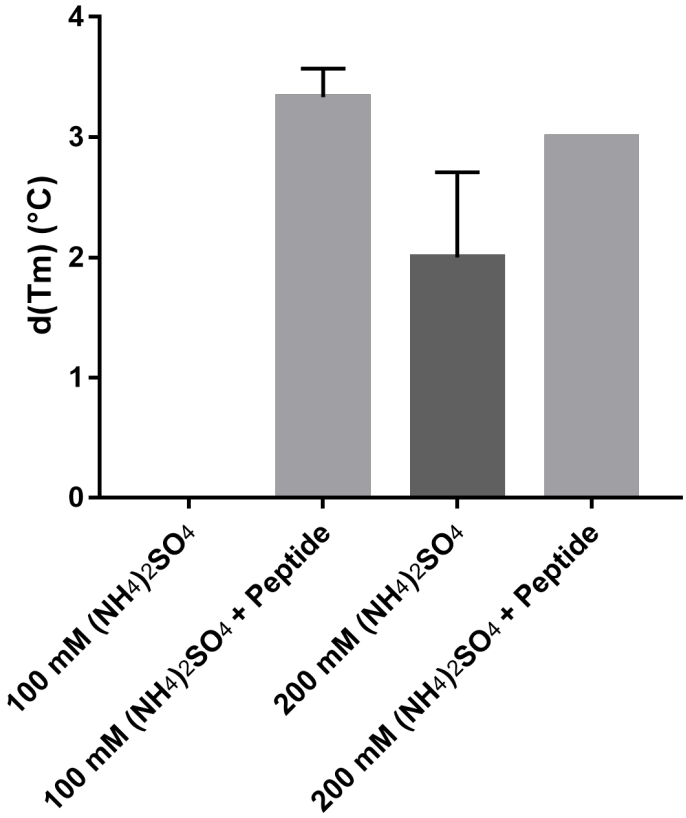

Figure 4.2.2. Binding of TBL1XR1$_{134-514}$ to MeCP2$_{285-309}$ in the presence of increasing concentrations of ammonium sulphate. TBL1XR1 in 100 mM ammonium sulphate (64.5˚C) was set as the reference. Error bars are one standard deviation

4.3 Rett mutations in MeCP2 abolish the interaction with TBL1X and TBL1XR1

As reported previously, Rett mutations in the NID abolish the interaction of MeCP2 with TBL1X in cell lysates (Lyst et al., 2013). To demonstrate that recombinant proteins behave similarly to the native mammalian proteins, initial experiments focussed on replicating these observations with the purified protein using pulldown assays, thermal denaturation assays and analytical gel filtration. The ability of purified TBL1X$_{176-527}$ to pull down wild type MeCP2 peptide as well as known RTT mutant peptides and non-binders was assayed (Figure 4.3.1).

Both wild-type $MeCP2_{285-313}$ and $MeCP2_{285-308}$ peptides were able to pull down TBL1X, though the R309X MeCP2 peptide, which has been shown to cause symptoms similar to RTT when expressed in knock-in mice, showed diminished binding (Moretti et al., 2006). None of the tested RTT mutants (K304E, K305R, and R306C) show any binding to TBL1X. $MeCP2_{285-313}$ containing a phosphorylated T308 is not able to pull down TBL1X either, consistent with previous observations that showed that when T308 is phosphorylated in response to neuronal activity, it abolishes the binding of MeCP2 to NCoR/SMRT (Ebert et al., 2013). This shows that the purified recombinant protein behaves similarly to the endogenous protein in *in vitro* assays.

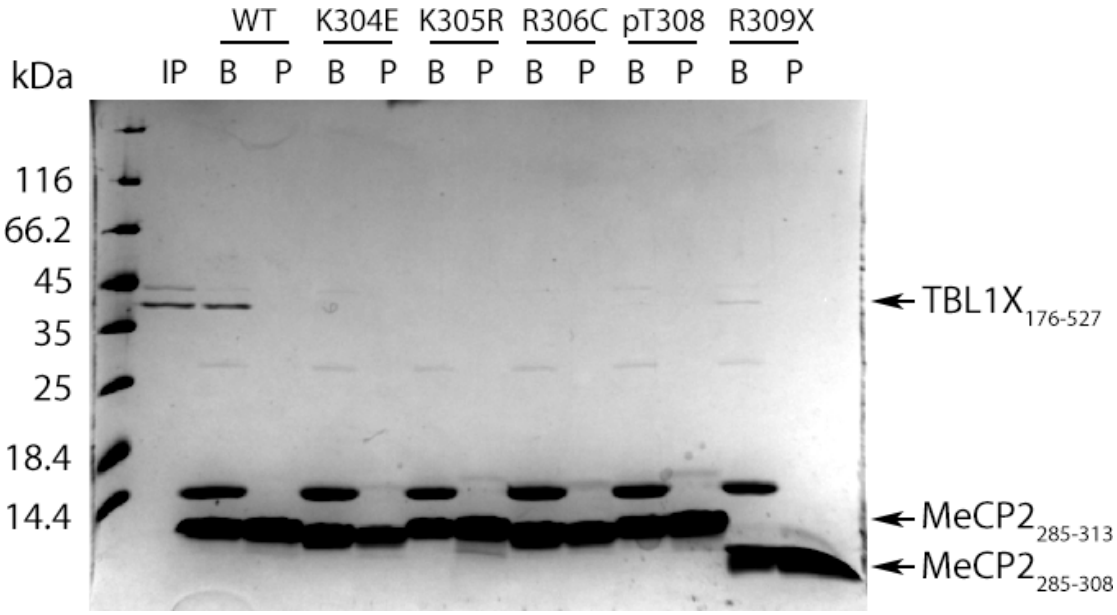

Figure 4.3.1. Pulldowns using $TBL1X_{176-527}$ and various MeCP2 peptides. IP – input. B – bound. P – peptide alone.

Binding of $TBL1X_{148-527}$ to $MeCP2_{285-313}$ peptide was also demonstrated using thermal denaturation assays. If it was possible to see stabilisation with wild-type MeCP2 peptide, but not with a peptide containing one of the RTT mutations, then the assay might be useful for screening small drug-like molecules that either inhibit or facilitate the interaction between TBL1X/TBL1XR1 and MeCP2. This could help treat either MeCP2 overexpression or RTT syndrome. I used $TBL1X_{148-527}$ purified from insect cells for the thermal denaturation assay. If wild-type MeCP2 NID peptide bound to TBL1X, but R306C mutant did not, I would observe a shift in the melting temperature with the wild-type peptide but not with the RTT R306C mutant. Consistent with previous observations I saw stabilisation of TBL1X when wild-type MeCP2 peptide was present with melting temperatures increasing from 60˚C with protein

alone to 64˚C in the presence of the peptide (Figure 4.3.2). With the R306C peptide, a 0.6˚C±0.23˚C shift was observed between the protein on its own, and the protein and the peptide. Although this could be interpreted as a small stabilisation and therefore a binding event, the difference is likely due to measurement error as fluorescence was sampled in 1˚C increments. Furthermore, the inability of MeCP2 protein containing R306C mutation to bind to TBL1X has been previously well-characterised. The assay also showed that TBL1X purified from insect cells behaves similarly to TBL1X purified from *E. coli*.

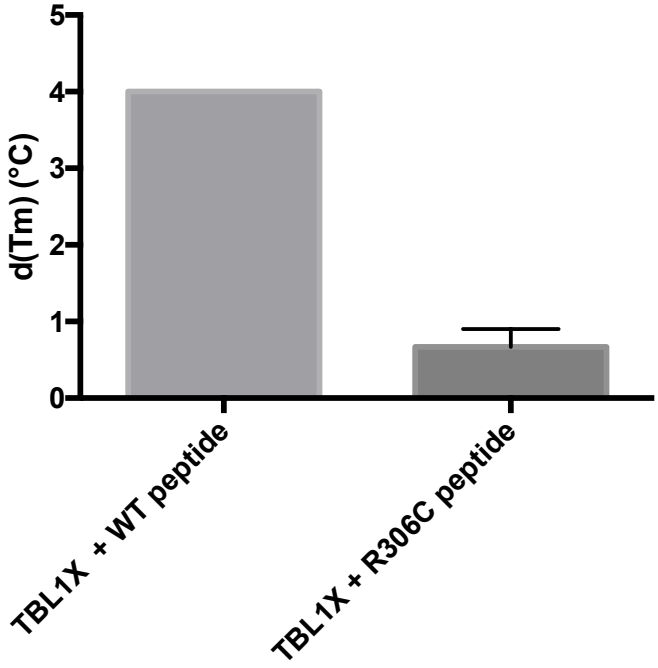

Figure 4.3.2. Thermal denaturation assay observing the stabilisation effect of $TBL1X_{148-527}$ by $MeCP2_{285-313}$ wild-type and R306C NID peptides. Melting temperature of $TBL1X_{148-527}$ (60˚C) was set as reference. Error bars are one standard deviation.

I also tested whether the TBL1XR1 construct equivalent to $TBL1X_{148-527}$, $TBL1XR1_{134-514}$, would bind to wild-type MeCP2 peptide using analytical gel filtration (Figure 4.3.3). As it was impossible to observe the peptide through absorbance at 280 nm as it lacks any aromatic residues, I measured the absorbance of TBL1XR1 instead. If binding of MeCP2 peptide to TBL1XR1 occurred, it should create a larger complex and thus a shift to lower elution volume should be observed, as the complex will have a larger hydrodynamic radius. As can be seen from Figure 4.3.3, the elution profiles of TBL1XR1 on its own and with R306C (orange and green lines, respectively) are identical save for the height of the peak which indicates slightly different amounts of proteins being injected onto the column. This difference likely arose due to pipetting errors when setting up the samples, as absorbance readings are volume

independent, depending only on the concentration. The identical peaks, however, mean that there has been no change in the hydrodynamic radius of the protein and thus there has been no binding event. Upon the addition of wild-type MeCP2$_{285-313}$ peptide to TBL1XR1 (purple line), there was a slight shift towards lower elution volume. This indicates that the protein has increased in its radius, and consequently the peptide is bound. It also showed that the protein and peptide form a stable complex in the context of this experiment. This suggested that co-crystallising TBL1XR1 and MeCP2 as a complex would be a good strategy, rather than soaking the peptide into existing crystals of TBL1XR1.

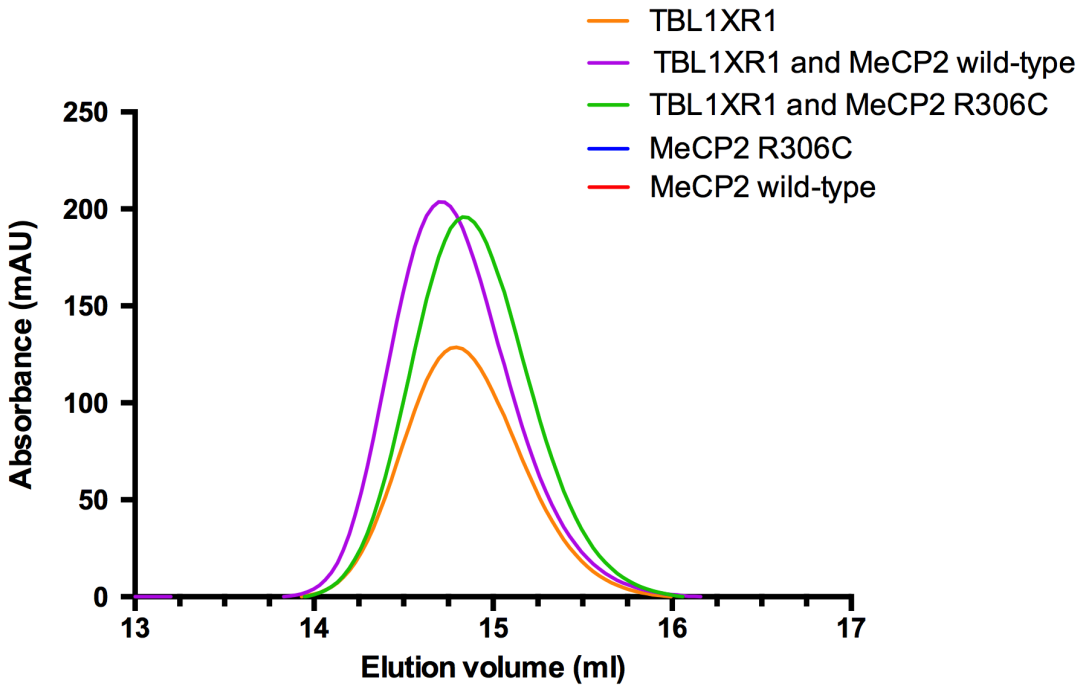

Figure 4.3.3. Analytical gel filtration profile using mouse TBL1XR1$_{134-514}$ construct, and either wild-type or R306C MeCP2$_{285-313}$ peptide. The wavelength at which absorbance was measured was 280 nm. The protein:peptide ratio was 1:1.2. Because the peptides lacked aromatic residues, they did not produce any signal at 280 nm.

4.4 Measuring the affinity of MeCP2 to TBL1XR1

To get an understanding of how strong the interaction between MeCP2 NID and TBL1XR1 is, two assays were performed. First a fluorescence anisotropy assay was used to measure the MeCP2 peptide binding to TBL1XR1. This technique involved using a fluorescent peptide, in this case N-terminal-fluorescein labelled MeCP2 peptides. The reaction mix containing the peptide and the protein is illuminated with a wavelength that is within the excitation spectrum of the fluorophore, while avoiding using excitation wavelengths that overlap with

the emission spectrum. After a certain time, the fluorophore will emit polarised light that is measured. If the peptide is not bound to a protein, its rotation will be fast and the emitted light will be depolarised. This will result in low anisotropy. However when the peptide is bound to a much larger protein, the tumbling rate will be much lower and the resulting anisotropy will be higher (Figure 4.4.1).

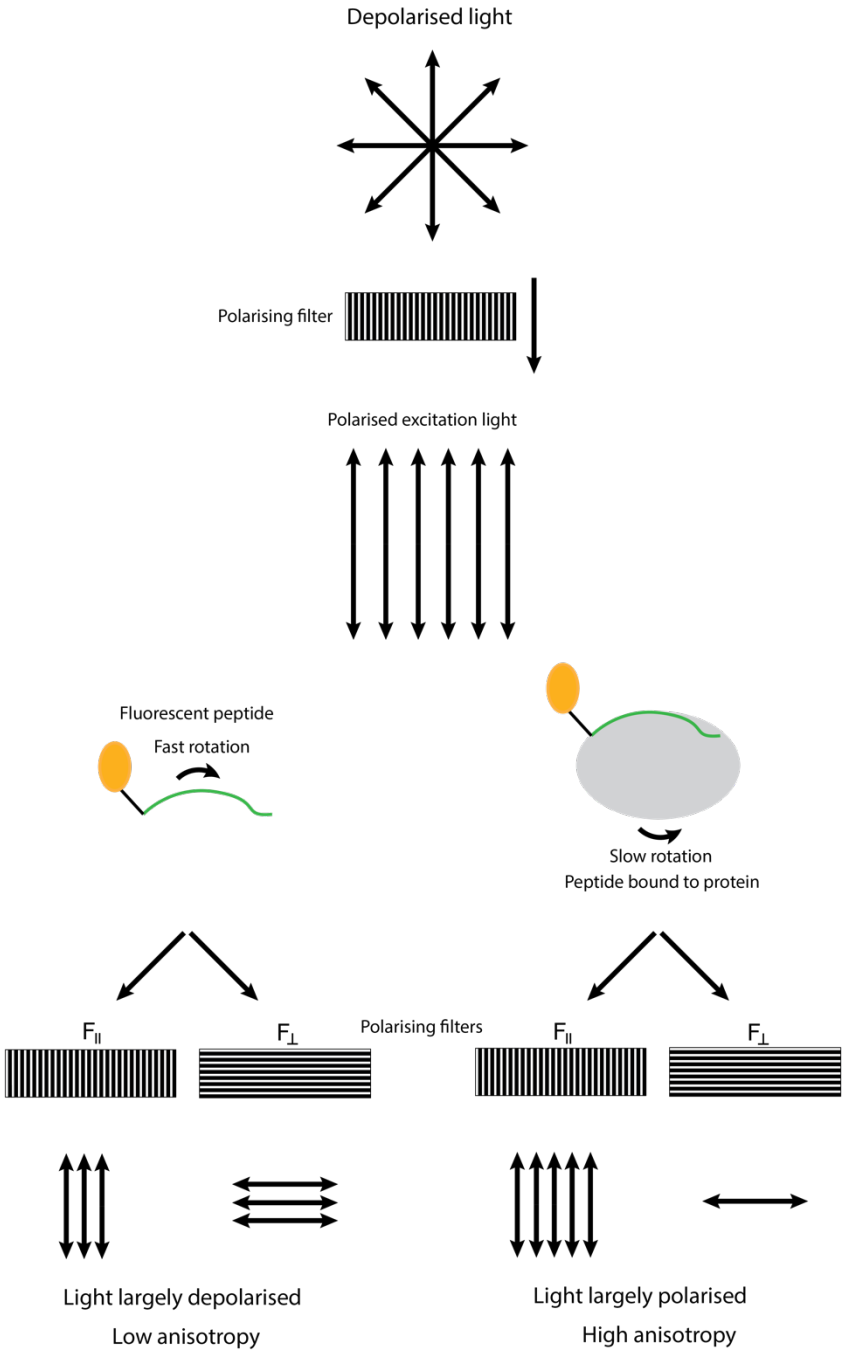

Figure 4.4.1. Principle of fluorescence anisotropy assay. Depolarised light passes through a polarising filter ($F_{||}$), which becomes polarised in the parallel direction. This light excites a fluorophore that is covalently linked to a small molecule. If the fluorophore-bearing small molecule

is not attached to a larger molecule, the tumbling rate of the molecule will be fast and the majority of the emitted light will be depolarised, measured as the intensity ratio of parallel to perpendicular ($F_{\perp}$) light (low anisotropy). However, if the fluorescently-tagged small molecule is attached to a larger molecule, the tumbling rate will decrease and the emitted light is largely polarised (high anisotropy).

Depending on the software used, the reported values from the experiment can either be parallel (F‖) and perpendicular (F⊥) fluorescence light intensities, polarisation (P), or anisotropy (r). To convert from measured perpendicular and parallel fluorescence intensity to anisotropy, equation 3.1 is used. To convert the measured light intensity to polarization, equation 3.2 is used (Lakowicz, 2006).

$$r = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + 2F_{\perp}} \quad \text{(Eq 3.1)}$$

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}} \quad \text{(Eq 3.2)}$$

Finally, to convert between anisotropy and polarization, equation 3.3 is used

$$r = \frac{2P}{3 - P} \quad \text{(Eq 3.3)}$$

Two peptides differing in the lengths of their N-termini, $MeCP2_{289\text{-}309}$ and $MeCP2_{298\text{-}309}$, were used to see whether regions beyond the RTT mutation hotspots were involved in binding. Initially a concentration series of these peptides were tested to determine the range of concentrations that would give a signal above the noise threshold while not saturating the instrument (Table 4.4.1 and Table 4.4.2). For the particular instrument used, the recorded intensity values should have been at least 20 as this was the noise limit of the machine.

Table 4.4.1. Measured values for $MeCP2_{289\text{-}309}$ N-terminal fluorescein tagged peptide.

Peptide concentration (nM)	Intensity	Average anisotropy	StDev of anisotropy
400	Saturated	N/A	N/A
200	332	0.062	0.004
100	119	0.074	0.002
50	49	0.083	0.002
25	20	0.088	0.002
12.5	8	0.081	0.004
6.25	4	0.079	0.006

Table 4.4.2. Measured values for MeCP2298-309 N-terminal fluorescein tagged peptide.

Concentration (nM)	Intensity	Average anisotropy	StDev of anisotropy
400	512	0.049	0.001
200	196	0.059	0.001
100	82	0.069	0.003
50	38	0.078	0.002
25	17	0.082	0.005
12.5	8	0.077	0.005
6.25	4	0.075	0.003

For $MeCP2_{289-309}$ and $MeCP2_{298-309}$ concentrations between 25-300 nM and 50-400 nM, respectively, produced sufficient signal. In MeCP2 and TBL1XR1 binding experiments, both 50 and 100 nM of $MeCP2_{289-309}$ and $MeCP2_{298-309}$ were tested again on the day of the experiment, as the maximum intensity values varied between experiments. Whichever concentration produced the lowest amount of signal above the threshold of 20 was then used in binding studies. For experiments shown in Figure 4.4.2 and Figure 4.4.3, the concentrations were 50 and 100 nM for $MeCP2_{289-309}$ and $MeCP2_{298-309}$, respectively. As can be seen from these figures, the $MeCP2_{298-309}$ peptide produced better signal than $MeCP2_{289-309}$, although the maximum anisotropy values attained were low in both cases and the data were noisy, especially at lower concentrations of TBL1XR1. In the case of $MeCP2_{289-309}$, the data were especially noisy and differences between binding to TBL1XR1 and BSA (negative control) were difficult to discern. At the same time, it could be seen from the binding curve of $MeCP2_{298-309}$ that there was no binding of the peptide to BSA, as otherwise there would be a non-saturable increase in anisotropy with increasing concentration of BSA. This showed that the observed binding of MeCP2 peptide to TBL1XR1 was a real effect. The anisotropy of $MeCP2_{289-309}$ free peptide at a concentration of 50 nM was 0.047±0.007 and the anisotropy of $MeCP2_{298-309}$ at a concentration of 100 nM was 0.049±0.005. These values were well within the expected range for a peptide of 1.8 kDa in size, which has a rotational correlation time smaller than the lifetime of fluorescein (4 ns) (Pope et al., 1999).

However, an issue with the assay was low levels of maximum anisotropy. For a complex of this size the expected limiting anisotropy is between 0.20 and 0.25 (Pope et al., 1999). It is possible that the lower values seen were due to 'propeller effect'. This happens when the

fluorophore is able to rotate independently of the molecule it is covalently linked to, reducing anisotropy. It is also conceivable that as the peptide is most likely unstructured, the flexible N-terminus of the peptide will increase the movement of the fluorophore and again decrease the anisotropy. While it was possible to analyse the data using non-linear regression and calculate a K_D from the data (Table 4.4.3), this value is likely to be erroneous as the anisotropy plateaued at low concentrations of TBL1XR1. This would skew the calculated K_D towards lower values.

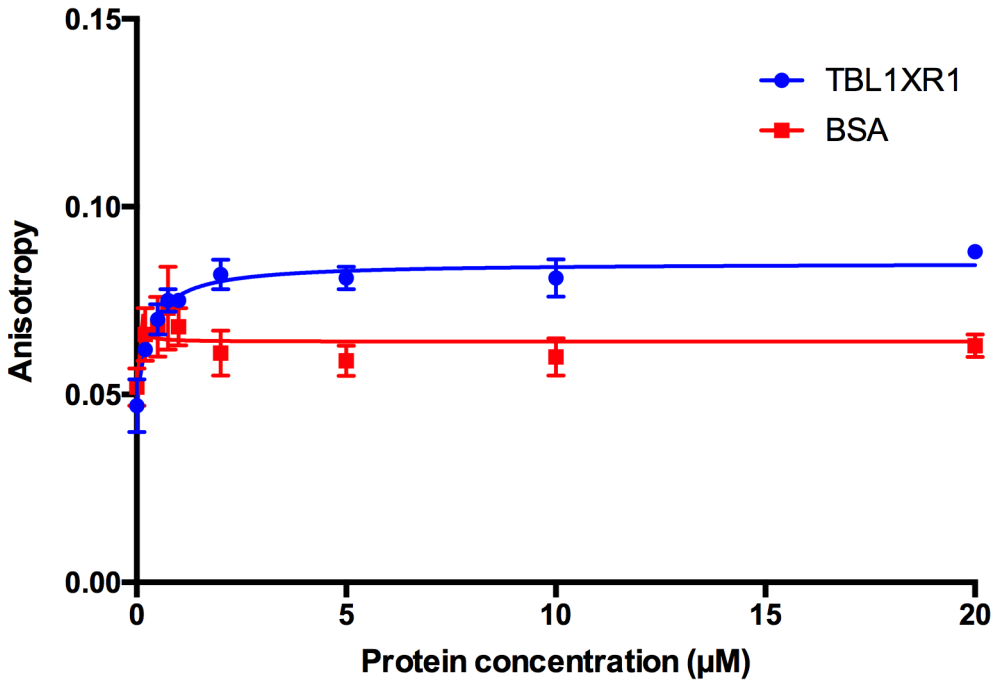

Figure 4.4.2. Fluorescence anisotropy assay measuring the binding of $MeCP2_{289-309}$ peptide to $TBL1XR1_{134-514}$. Concentration of MeCP2 peptide was 50 nM. BSA was included as a negative control. All experiments were done in triplicate. Binding curves were fitted using non-linear regression. Error bars are one standard deviation.

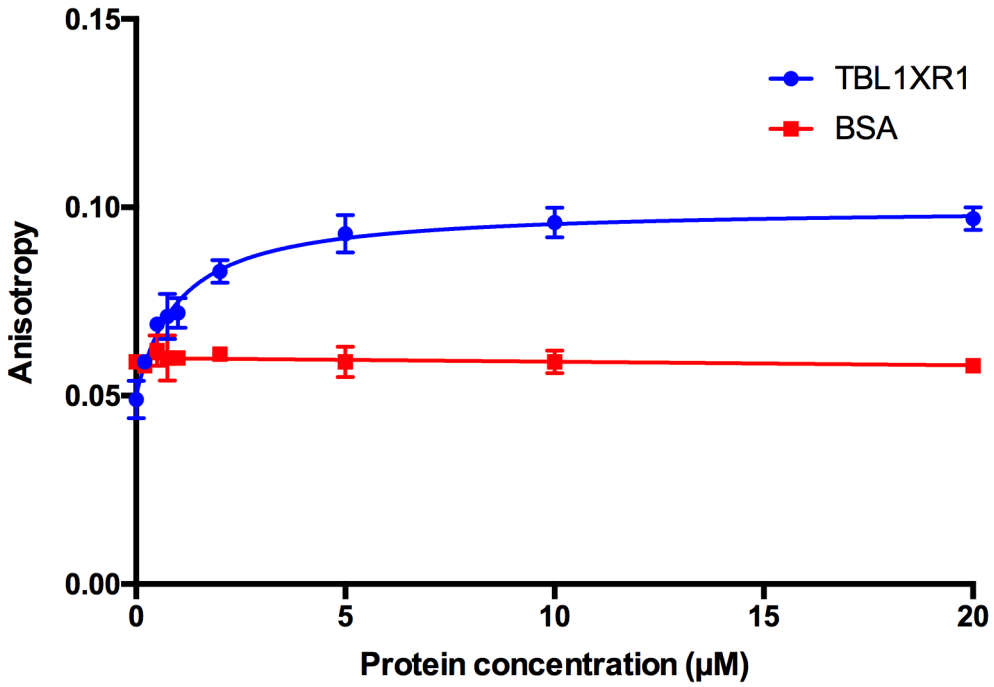

Figure 4.4.3. Fluorescence anisotropy assay measuring the binding of MeCP2$_{298-309}$ peptide to TBL1XR1$_{134-514}$. Concentration of MeCP2 peptide was 100 nM. BSA was included as a negative control. All experiments were done in triplicate. Binding curve was calculated using non-linear regression. Error bars are one standard deviation.

Table 4.4.3. Analysis of results obtained from fluorescence anisotropy experiments shown in Figure 4.4.2 and Figure 4.4.3 using non-linear regression.

Protein	K_D (μM)	Standard Deviation (μM)
MeCP2$_{289-309}$	0.29	0.06
MeCP2$_{298-309}$	0.96	0.17

To see how the peptide behaves in solution and whether the measured anisotropy values change depending on how long the solution has been left to reach equilibrium, I set up another assay with fresh aliquots of MeCP2$_{298-309}$ peptide and fresh preparation of TBL1XR1$_{134-514}$. MeCP2$_{298-309}$ peptide was used because the aim was to keep the flexible non-binding regions as short as possible. A higher concentration of TBL1XR1 was also used, to make sure that the binding was close to saturation. On this occasion the anisotropy values spanned a much wider range and neared the expected value of 0.20. There was not much change in the values recorded at 0 min and at 60 min. This showed that shorter or longer incubation times do not adversely affect the measurements. Averaging the data over the

three time points gave a K_D value of 20.6 ± 1.3 µM. Individual time point analysis results are given in Table 4.4.4.

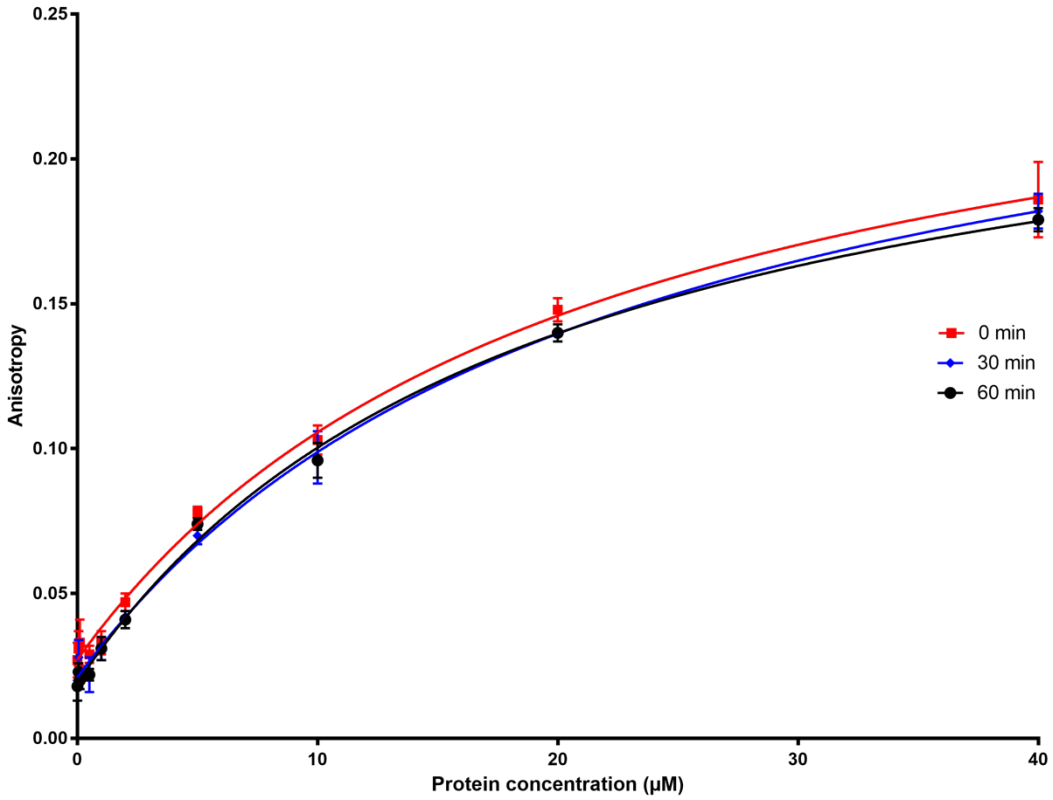

Figure 4.4.4. MeCP2$_{298-309}$ binding to TBL1XR1$_{134-514}$. Measurements were taken at 0, 30, and 60 min.

Table 4.4.4. Analysis of results obtained from fluorescence anisotropy experiments shown in Figure 4.4.4 using non-linear regression.

Time point	K_D (µM)	Standard Deviation (µM)
0 min	20.9	2.7
30 min	22.0	2.3
60 min	18.8	1.6

It wasn't immediately clear what had caused the low anisotropy values recorded in the first experiments. However, as the two experiments gave K_D values that differed from each other by a factor of 20, a parallel method was required to verify the binding constant obtained from the second experiment. This was done by using surface plasmon resonance (SPR) assays. The principle of the assay is to measure binding of ligands to receptors (or vice versa) immobilised on a sensor chip by measuring a dip in the intensity of the reflected light. A beam of light is passed through a prism under total internal reflection (TIR) conditions. This beam of light hits

the sensor chip, which is coated in a gold film. Under certain conditions (depending on the angle or the wavelength of the incident light), the photons can be absorbed by the electrons in the free electron cloud of the gold atoms, giving rise to surface plasmons, which are electromagnetic waves. This absorption of photons and resulting drop in reflectance can be detected in the reflected light beam hitting a detector (Figure 4.4.5A). As ligands bind to immobilised receptors, the local environment, and consequently the resonance angle (or wavelength, depending on the setup) changes, which is, again, detected by the sensor (Figure 4.4.5B).

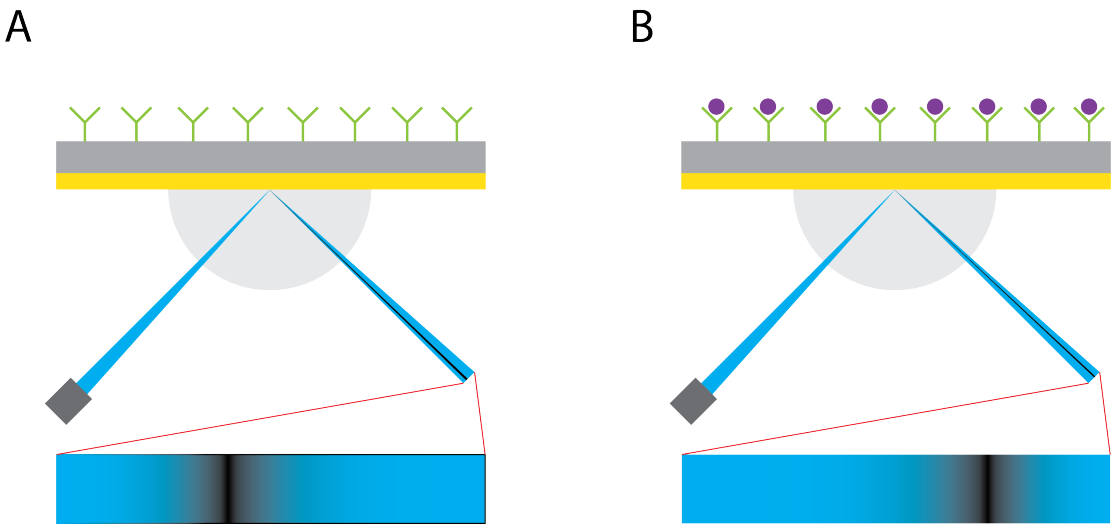

Figure 4.4.5. Principle of surface plasmon resonance. A) At a certain angle of incident light, some of the photons hitting the gold surface of the SPR chip are absorbed by the electron cloud of the gold atoms. These convert electrons into surface plasmons, which results a drop in the intensity of the reflected light at a certain angle of reflection (black band). B) As other molecules bind, they change the local environment and thus the resonance angle, leading to a shift in the intensity drop. This is detected by the instrument and displayed in the software.

Three peptides were used - MeCP2$_{285-313}$ wild-type and R306C, and wild-type MeCP2$_{285-309}$. These peptides were chosen as MeCP2$_{285-309}$ is the classically defined NID and sufficient for binding of MeCP2 to TBL1XR1, while MeCP2$_{285-313}$ might display enhanced binding. MeCP2$_{285-313}$ R306C was chosen as a negative control.

Initially there were difficulties with the assay as TBL1XR1$_{134-514}$ had a strong affinity for the carboxymethylcellulose matrix used. The protein would not dissociate from the chip unless it was washed with 0.5% SDS. Numerous attempts to use detergents and commercial mixes to reduce non-specific binding had no effect. Altering the salt concentration did not have the desired effect either. Supplementing the protein sample with 100 U/ml of heparin effectively

solved the non-specific binding issue. It introduced a problem with buffer mismatch and subsequent large injection start and stop peaks, but as the kinetics of the interaction were too fast to be resolved anyway, affinities could only be calculated from steady-state.

Figure 4.4.6 shows the binding curves of the three peptides to $TBL1XR1_{134-514}$. The dimensionless response units (RU) correspond to change of mass on the chip surface. Because more $MeCP2_{285-309}$ than $MeCP2_{285-313}$ was attached to the chip (as measured in change of RU), the maximum response of binding was also higher. The measured dissociation constants were 9.5 ± 0.5 µM and 12.9 ± 0.8 µM for $MeCP2_{285-313}$ and $MeCP2_{285-309}$ respectively. This implied that while $MeCP2_{285-309}$ was sufficient for the binding to TBL1XR1, amino acids 310-313 may be involved in enhancing this interaction. These were consistent with the measured K_D values using the fluorescence anisotropy assay. As expected, $MeCP2_{285-313}$ R306C did not bind to any measurable extent. Figure 4.4.7, Figure 4.4.8, and Figure 4.4.9 show the raw data obtained from the experiments. Because of the slight delay of buffer flow between flow-cells, the $MeCP2_{285-309}$ and $MeCP2_{285-313}$ R306C (flow-cells 3 and 4, respectively) show more pronounced buffer mismatch spikes than $MeCP2_{285-313}$ which was attached to flow-cell 2. Fluctuating RU values were observed during late injection phase (Figure 4.4.8, between 20 and 30 s), though it did not pose a problem for data analysis as steady state analysis could be done on earlier time points.

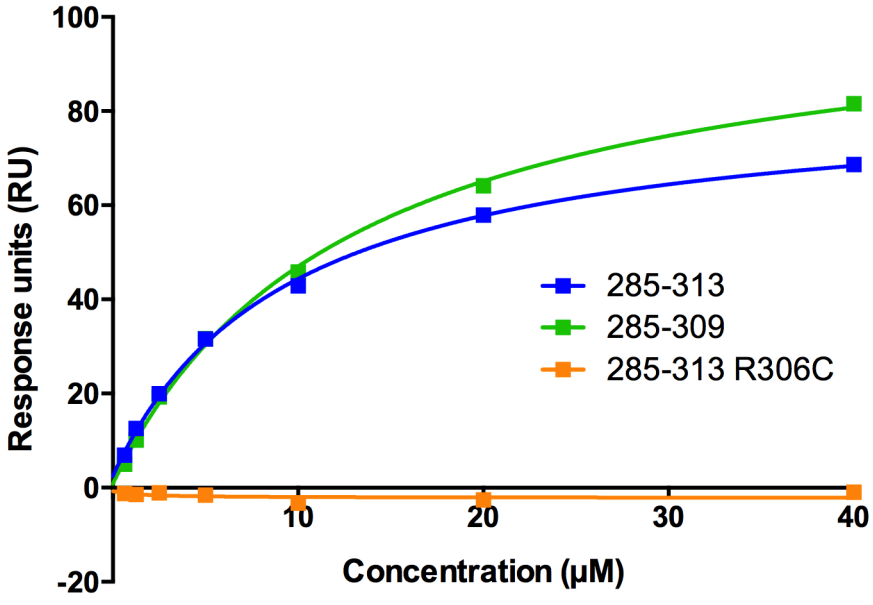

Figure 4.4.6. Binding curves of $MeCP2_{285-313}$, $MeCP2_{285-309}$, and $MeCP2_{285-313}$ R306C peptides to wild-type $TBL1XR1_{134-514}$. R306C peptide was included as a negative control. Data was fit using non-linear regression to Langmuir isotherm.

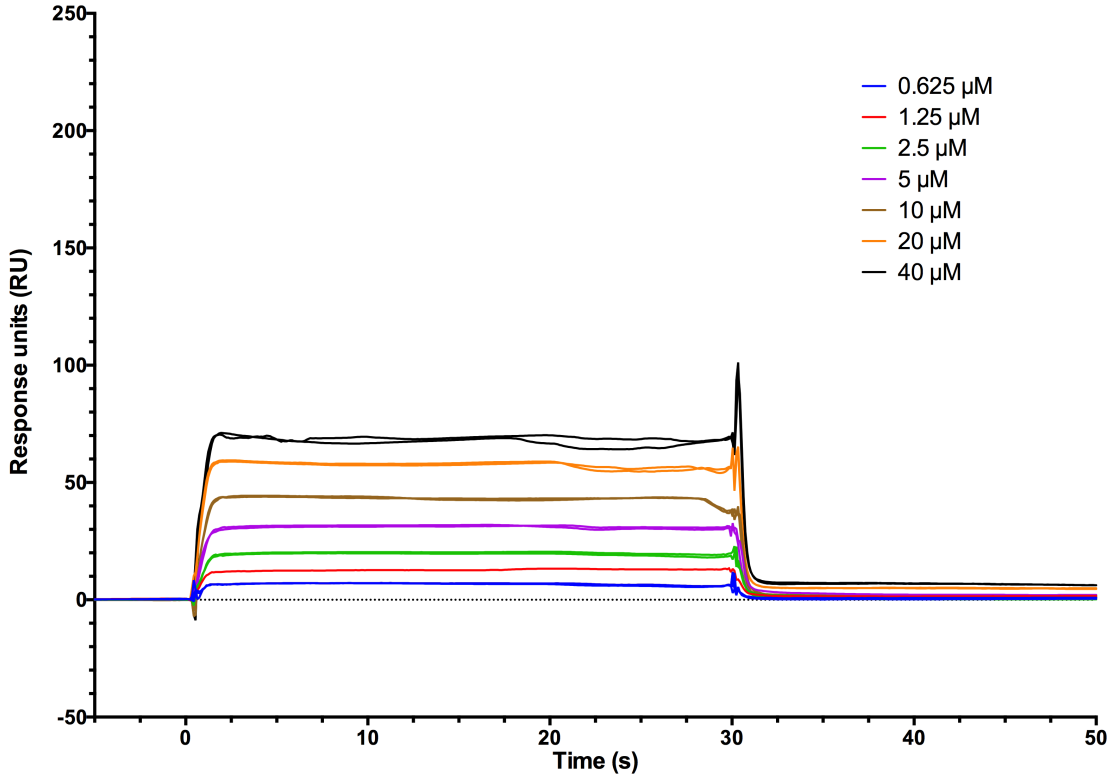

Figure 4.4.7. Binding of a concentration series of TBL1XR1$_{134-514}$ to MeCP2$_{285-313}$ peptide

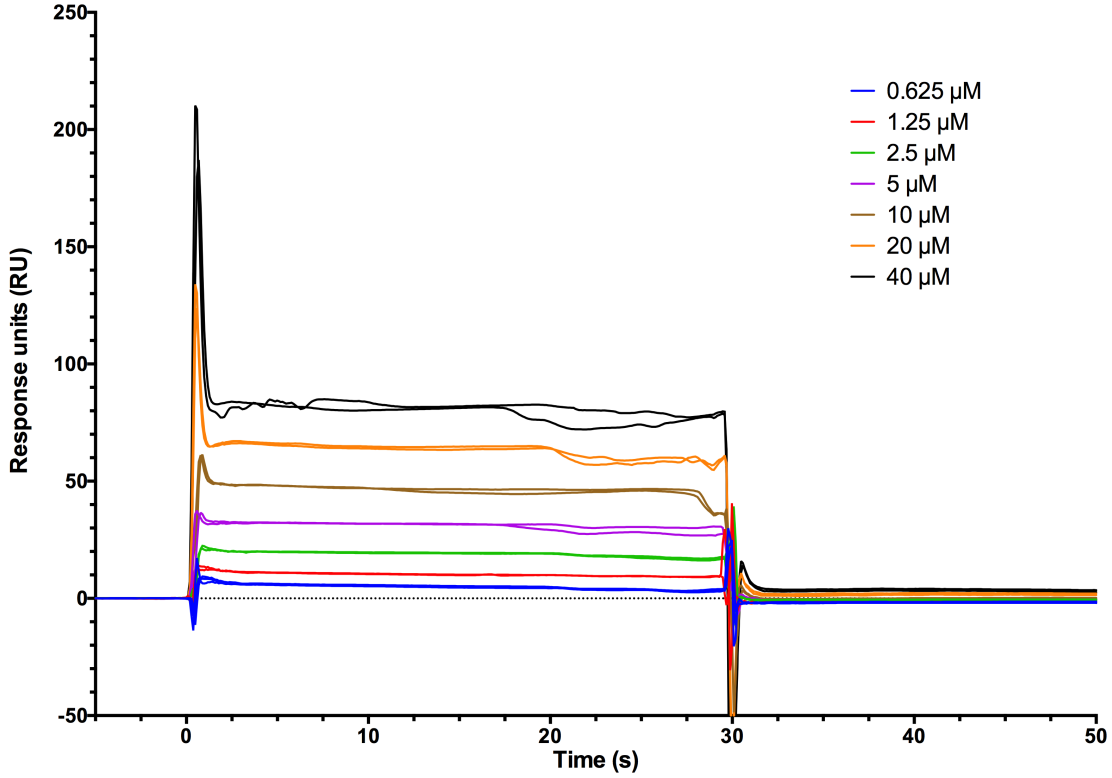

Figure 4.4.8. Binding of a concentration series of TBL1XR1$_{134-514}$ to MeCP2$_{285-309}$ peptide

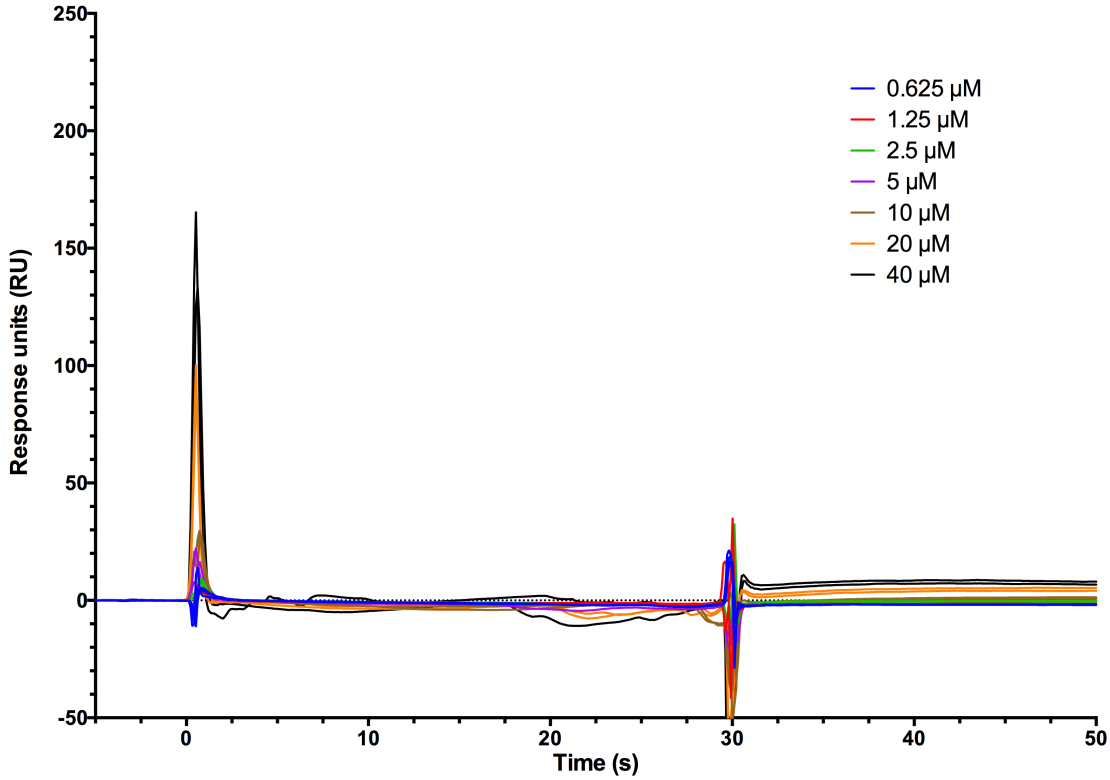

Figure 4.4.9. Binding of a concentration series of TBL1XR1$_{134-514}$ to MeCP2$_{285-313}$ R306C peptide

4.5 Discussion

As the surface of TBL1XR1 where MeCP2 was thought to bind had a strong electronegative character, I wanted to test whether the binding of MeCP2 had an electrostatic component. To this end I tested different concentrations of NaCl and AmSO4. When comparing the thermal stability of TBL1XR1 alone and in solution with MeCP2$_{285-309}$ peptide, it emerged that the stabilisation effect due to the peptide binding to TBL1XR1$_{134-514}$ was not present at around 400 mM NaCl and at 200 mM ammonium sulphate. While thermal denaturation assays don't measure the binding directly, it indirectly indicated that the binding of the MeCP2 peptide was lost in those conditions. This was in agreement with IP-mass spectrometry experiments done in the Bird lab where components of the NCoR/SMRT complex (i.e. TBL1X and TBL1XR1) were only seen interacting with MeCP2 near physiological salt concentrations.

Purified recombinant TBL1X and TBL1XR1 were tested for their ability to bind to wild-type MeCP2 peptides and to known RTT mutants. Initial pulldowns showed that TBL1X$_{176-527}$ binds to wild-type MeCP2$_{285-313}$ peptide, and to a lesser extent to MeCP2$_{285-308}$ peptide, however no binding to was observed to MeCP2$_{285-313}$ K304E, K305R, and R306C RTT peptides, or to

phosphorylated T308 which is a residue involved in neuronal activity-regulated dissociation of MeCP2 from TBL1X/TBL1XR1. This was confirmed using TDA where I showed that while $MeCP2_{285-313}$ stabilises $TBL1X_{148-527}$, no stabilization is observed when I use the same peptide with an R306C mutation. Semi-analytical gel-filtration using $MeCP2_{285-313}$ and $TBL1XR1_{134-514}$ further confirmed this by producing no change in the elution volume when $TBL1XR1_{134-514}$ was mixed with $MeCP2_{285-313}$ R306C, however upon the addition of wild-type $MeCP2_{285-313}$ a larger species was observed, indicating that binding had occurred.

Finally, I measured the affinity of this interaction using fluorescence anisotropy assays and surface plasmon resonance assays. Initially there were some problems with the fluorescence signal due to a potentially unrestricted fluorophore causing a decay in anisotropy. However, using a shorter peptide ($MeCP2_{298-309}$) with $TBL1XR1_{134-514}$ produced measurable responses from which dissociation constant value of 20.6 ± 1.3 µM could be calculated. This value was consistent with surface plasmon resonance assays using $MeCP2_{285-313}$ and $MeCP2_{285-309}$ peptides, from which dissociation constant values of 9.5 ± 0.5 µM and 12.9 ± 0.8 µM could be calculated. The R306C peptide showed no binding to TBL1XR1, as expected. Although SPR and FA are fundamentally different methods, with the first using immobilized molecules and the second being a solution-based assay, they agree well in their estimates. Taken together, these results suggest that the dissociation constant of the interaction between MeCP2 and TBL1XR1 is around 10-20 µM.

5 Structural and biochemical characterisation of TBL1XR1 and MeCP2 interaction

5.1 Introduction

An NMR solution structure of the MeCP2 MBD was published in 1999 (Wakefield et al., 1999), and a crystal structure of the MeCP2 MBD bound to methylated DNA was published in 2008 (Ho et al., 2008). The latter structure showed that MeCP2 MBD recognises DNA methylation through structurally conserved waters in the major groove. However, as of July 2016 there are no other structures available for the other parts of MeCP2. In 2014 a crystal structure of the WD40 domain of human TBL1XR1 was deposited in the PDB (PDB 4LG9) (Xu et al., 2013). To obtain co-crystals of MeCP2 and TBL1XR1, two different approaches could have been taken. MeCP2 and TBL1XR1 could have been co-crystallised by either co-purifying or mixing the two proteins together to form a complex, followed by crystallisation. Alternatively, MeCP2 could have been soaked into crystals of TBL1XR1. Analysing the human TBL1XR1 crystal structure and conditions showed that the conditions included high salt (2 M AmSO4), which would have abolished the binding of MeCP2 to TBL1XR1 as revealed by previous TDA experiments. Furthermore, the observed crystal packing would have been unable to accommodate the MeCP2 peptide, as the central channel of the WD40 domain makes contacts with a loop from a TBL1XR1 molecule in an adjacent asymmetric unit, effectively blocking E364, a residue that we know from unpublished data to be important in the interaction (Figure 5.1.1). This would have made soaking in MeCP2 peptide impossible. It was therefore necessary to come up with novel crystallisation conditions that would allow for the co-crystallisation of MeCP2 with TBL1XR1. This could have been achieved by crystallising the mouse TBL1XR1 in different conditions that would have been able to accommodate the MeCP2 peptide. However, I chose to co-crystallise the complex instead.

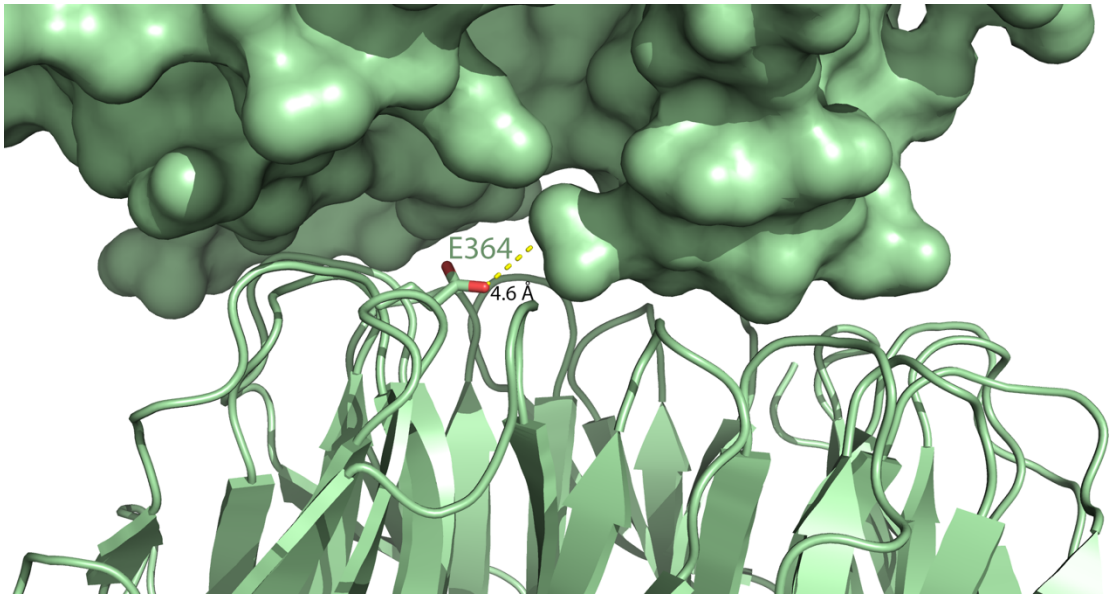

Figure 5.1.1. Crystal packing seen in human TBL1XR1 crystal (PDB: 4LG9). The distance from E364 to the next subunit is 4.6 Å.

Understanding the interaction between TBL1XR1 and MeCP2 in atomic detail should give insight into how mutations in the NID lead to RTT. By understanding the binding site in detail, as well as which residues are critical for this interaction to occur, it may be possible to use rational design for the development of small drug-like molecules. This could lead to therapeutic treatment of MeCP2 overexpression or RTT syndrome by either inhibiting or facilitating the interaction between MeCP2 and TBL1XR1.

In this chapter the first structural data about the interaction between TBL1XR1 and MeCP2 NID is reported. The contribution of each of the NID residues seen mutated in RTT, as well as reasons why these mutations would lead to loss of binding are discussed in detail. In addition, it is shown that reciprocal mutations in the TBL1XR1 binding site that do not cause destabilisation of the WD40 fold can equally abolish the binding of MeCP2. Finally, the affinity between TBL1XR1 binding site mutants and two MeCP2 peptides, $MeCP2_{285-309}$ and $MeCP_{285-313}$, are investigated using surface plasmon resonance.

5.2 Sparse-matrix screening

While previous experiments showed that MeCP2 is unable to bind to TBL1XR1 in the presence of high salt, some of the crystal trays were set up prior to those experiments, and a few screens tested (such as AmSO4 screen) contained conditions that, in hindsight, would have prevented co-crystallisation. Multiple commercial screens were tested in 96-well plate

format, including INDEX, JCSG+, Anions and AmSO4 suite, as well as Magic I and II, and Complex I and II screens obtained from Max Planck Institute for Biochemistry, Martinsried. Different protein concentrations and drop sizes were tested, although these did not produce any hits. A few plates were set up at room temperature and then transferred to 4˚C, however no crystals were obtained this way. On a few occasions crystals were observed in the screens, however when sent to synchrotron facility they were found to be salt crystals. Overall, while sparse-matrix screening failed to produce crystals, it provided information for optimization as often the wells had gelatinous precipitation or spherulites, which indicated that the protein was supersaturated and forming nucleation sites but the amount of protein was probably too high, which led to rapid nucleation.

5.3 Optimisation of crystallisation conditions

Conditions that showed gelatinous precipitation or spherulites often had PEG as the common precipitation agent. It was decided to test a range of PEGs at different concentrations using 50 mM MOPS pH 7.5 as the buffer to see if it was possible to obtain any crystals. The initial test was carried out in a 24-well hanging drop plate using TBL1XR1$_{134-514}$ at a concentration of 12.6 mg/ml mixed with MeCP2$_{285-309}$ NID peptide in a 1:1.2 ratio of protein:peptide, and observing the effect the PEGs (Table 5.3.1).

Table 5.3.1. Hanging drop results for co-crystallisation of TBL1XR1$_{134-514}$ mixed with MeCP2$_{285-309}$ in 1:1.2 protein:peptide ratio. X denotes immediate brown precipitation (denatured protein), while † denotes overnight light amorphous precipitation. Empty cells indicate clear drops. * denotes microneedles that grew after five days.

	PEG 400	PEG 550	PEG 2000	PEG 3350	PEG 5000
10%					X
15%	†	†			X
20%	X	X	†	*	X
25%	X	X	†	†	X

As PEG 5000 caused heavy precipitation of the protein at concentrations 10% and above, this PEG was not used further, and focus was switched to a range of lower molecular weight PEGs. To assess the effect of these PEGs, another 24-well plate was set up where the concentration of PEGs was varied between the conditions that produced clear and precipitated drops. PEG

1000 was also included to test its ability as a precipitation agent at similar concentrations (Table 5.3.2). While nothing appeared in the drops during the next two days, after five days some of the conditions in these two plates produced microneedles and spherulites (Figure 5.3.1).

Table 5.3.2. Hanging drop results for co-crystallisation of TBL1XR1$_{134-514}$ (12.6 mg/ml final concentration) mixed with MeCP2$_{285-309}$ in 1:1.2 protein:peptide ratio. * denotes microneedles, ° denotes spherulites. Empty cells contained light amorphous precipitation.

Concentration	PEG 550	PEG 1000	PEG 2000	Concentration	PEG 3350
13%				**21%**	*
15%				**23%**	*
17%			°		
19%			°		

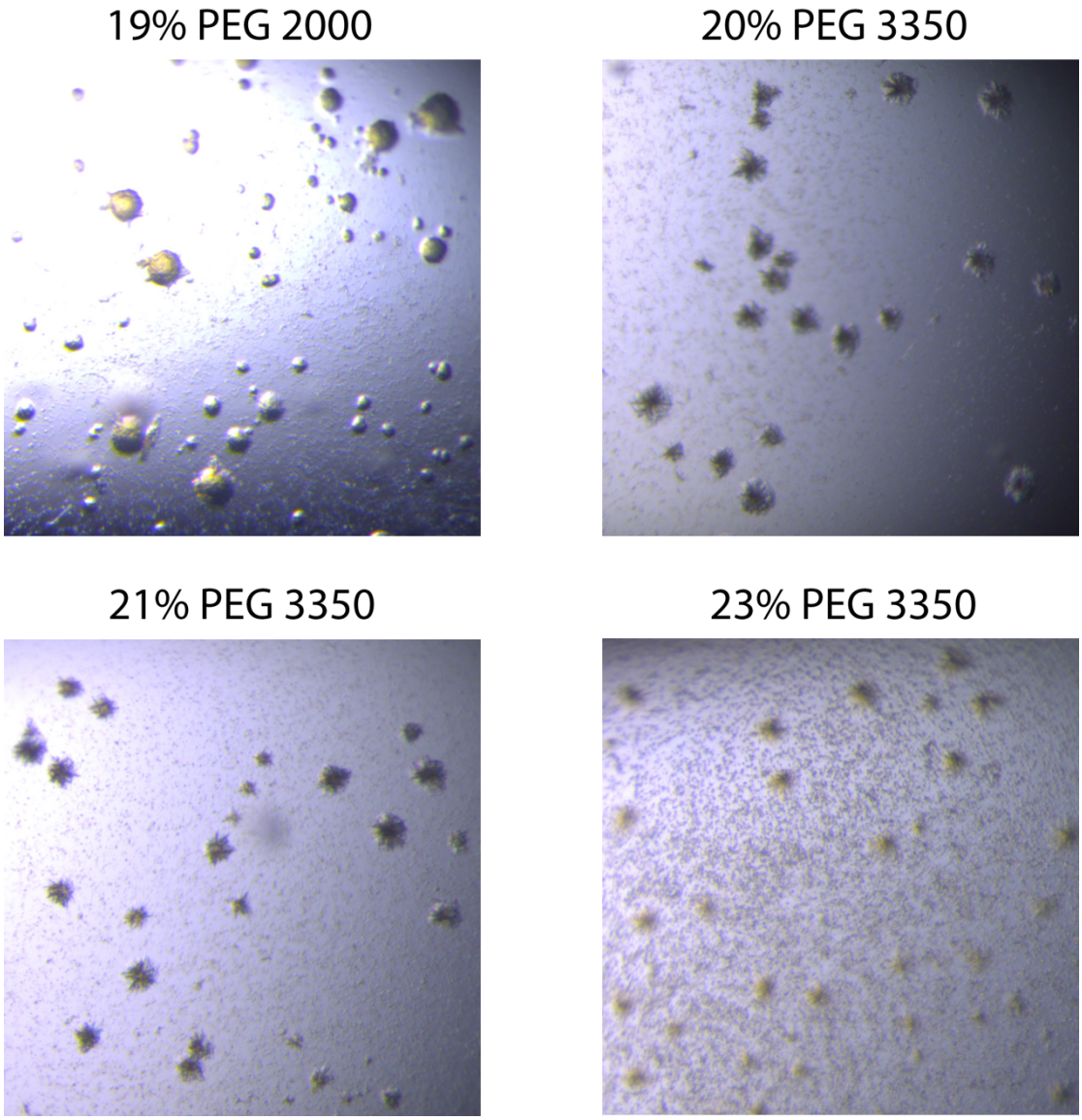

Figure 5.3.1. Microneedles and spherulites of TBL1XR1$_{134-514}$ and MeCP2$_{285-309}$ obtained in 50 mM MOPS pH 7.5.

The optimal concentration of PEG 3350 was between 17% and 21%. Conditions around these values were optimised to obtain co-crystals of TBL1XR1 and MeCP2. The first 24-well plate probed the effect of pH on the growth of crystals. Using MOPS as a buffer, the pH was varied between 6.5 and 8.0 in 0.5 unit increments (Table 5.3.3). However, after a week none of the conditions had grown crystals. Even the replicated drops with 19% and 21% PEG 3350 at pH 7.5 only had precipitation.

Table 5.3.3. Hanging drop results for co-crystallisation of TBL1XR1$_{134-514}$ (12.6 mg/ml final concentration) mixed with MeCP2$_{285-309}$. Empty cells indicate light amorphous precipitation.

PEG 3350	MOPS pH 6.5	MOPS pH 7.0	MOPS pH 7.5	MOPS pH 8.0
17%				
19%				
21%				

A number of plates were set up in the following weeks. Because so far the best results had been obtained with PEG 2000 and PEG 3350, the plates had just these two PEGs in concentrations between 17% and 23%. Due to the stochastic nature of crystallisation, conditions that had produced microneedles were set up each time as well. Unfortunately, none of the plates produced anything other than precipitation. Lowering the TBL1XR1$_{134-514}$ final concentration produced either clear drops or drops with slight precipitation. As there was a drop with microneedles, they were used those as a stock for crystal seeding in 21% and 23% PEG 2000 and 3350 at ratios ranging from 1:10 to 1:100, but no crystal growth was observed. Instead again the drops showed precipitation.

Based on vacuum electrostatics calculated in PyMol, some of the areas on TBL1XR1 were highly electronegative. Addition of spermidine in low concentrations was therefore tried. It was hoped that spermidine would neutralise some of the more electronegative regions that might cause electrostatic repulsion and inhibit crystal packing. Addition of 20 mM and 50 mM spermidine to crystallisation conditions containing 50 mM MOPS pH 7.5 and varying concentrations of PEG 3350 (14%-24%, 2% increments) produced gelatinous precipitate, which indicated that it could prove helpful in crystallisation attempts. Unfortunately, subsequent trials varying buffers and their concentrations, pH values, or the PEGs used did not produce anything similar to the microneedles previously observed.

During the optimisation phase the older plates were observed weekly. At one point stacked crystal plates were seen in wells that contained 100 mM MOPS pH 7.5 and 16%, 18% and 20% PEG 3350 with final TBL1XR1 concentration of 10 mg/ml (Figure 5.3.2). These conditions were very similar to those that produced microneedles, with the exception that the concentration of MOPS buffer was 100 mM, not 50 mM, and TBL1XR1 concentration was lower than what was initially tested. These conditions were observed for a week to see whether any new single crystals would appear, or whether any of the old crystals disappeared, yet there was no change after the initial appearance of the crystals.

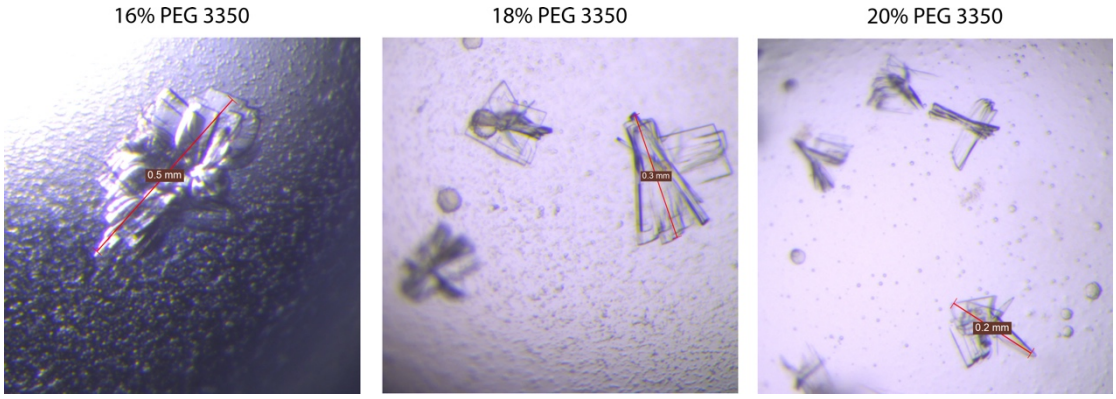

Figure 5.3.2. Stacked plates of TBL1XR1$_{134-514}$ and MeCP2$_{285-309}$ crystals obtained in 100 mM MOPS pH 7.5.

As there was no change, the stacked plates were carefully dissected to produce single crystals, flash-cooled in mother liquor with 30% glycerol as cryoprotectant using liquid N_2, and sent off to Diamond Light Source for data collection. While many of the crystals did not produce good diffraction data, or did not diffract at all, one of the crystals diffracted to a resolution of 2.5 Å. The data from that crystal was processed, after which molecular replacement was carried out using human TBL1XR1 (PDB ID: 4LG9) as a template. The obtained model was rebuilt using COOT and refined in Phenix. Data relating to the refined structure has been shown in Table 5.3.4.

Table 5.3.4. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses. Output from Phenix.

	TBL1XR1$_{134-514}$ and MeCP2$_{285-309}$
Wavelength	0.976250 Å
Resolution range	49.6 - 2.5 (2.6 - 2.5)
Space group	*P* 1
Unit cell (a,b,c,α,β,γ)	42.6 58.6 155.1 97.6 91.2 90.2
Total reflections	118363 (10840)
Unique reflections	49872 (4583)
Multiplicity	2.4 (2.4)
Completeness (%)	0.98 (0.88)
Mean I/sigma(I)	9.82 (2.25)
Wilson B-factor	33.30
R-merge	0.078 (0.429)
R-meas	0.101 (0.555)
CC1/2	0.993 (0.726)
CC*	0.998 (0.917)
Reflections used in refinement	50630 (4583)
Reflections used for R-free	2511 (246)
R-work	0.207 (0.264)
R-free	0.243 (0.294)
CC(work)	0.938 (0.824)

CC(free)	0.911 (0.768)
Number of non-hydrogen atoms	10343
Macromolecules	9569
Peptides	155
Waters	511
Glycerol	108
Protein residues	1384
RMS(bonds)	0.005
RMS(angles)	0.67
Ramachandran favoured (%)	95.6
Ramachandran allowed (%)	4
Ramachandran outliers (%)	0.4
Rotamer outliers (%)	0.81
Clashscore	3.11
Average B-factor	46.86
macromolecules	47.20
ligands	54.29
solvent	38.86

5.4 MeCP2 binds to the top face of the TBL1XR1 WD40 domain

There were four molecules of TBL1XR1$_{134\text{-}514}$ present in the asymmetric unit (chain A-D) and two of these molecules (A and B) had MeCP2 bound (chain E and F, respectively). These two TBL1XR1 molecules (A and B) were the best defined with side chains clearly visible for most residues, spanning residues 155 to 513, and 155 to 514 in chains A and B, respectively. The remaining two molecules with no MeCP2 NID peptide bound had poor electron density, especially for the side-chains, many of which are trimmed to the Cα atom as there was no supporting density to model in the rest of the side-chain (Figure 5.4.1). In some regions the main chain was removed as well, as a decision was made not to model anything that wasn't supported by electron density. Consequently, while these two molecules spanned residues 158-513 and 159-512, many of the connecting loops between adjacent beta-strands were missing.

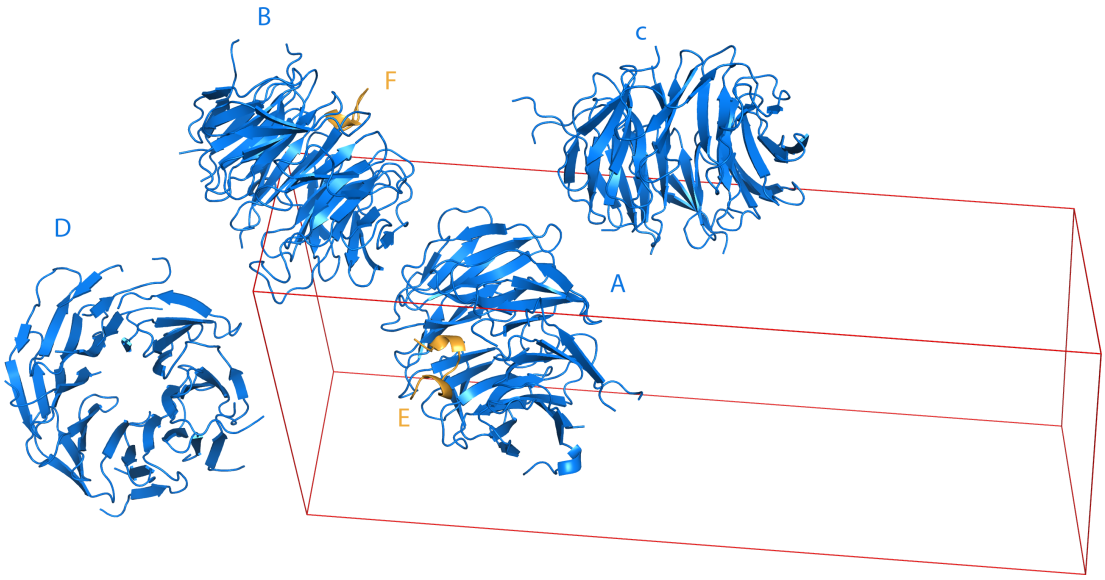

Figure 5.4.1. Asymmetric unit of TBL1XR1$_{134-514}$ (blue) and MeCP2$_{285-309}$ (orange) co-crystal. There were 4 molecules of TBL1XR1$_{134-514}$ (chains A-D), and two molecules of MeCP2$_{285-309}$ (chains E and F) present in the asymmetric unit. Unit cell has been shown in red.

Disulphide bridges linking the molecules in the asymmetric unit with their neighbours in the adjacent unit cells (marked with an apostrophe) were observed between C214^{TBL1XR1} in chain A and chain C', chain B and chain D', chain C and chain A', and chain D and chain B' (Figure 5.4.2). Attempts were made to model these linkages into the PDB file, however creating a link between molecules in different asymmetric cells was unsuccessful, so in the PDB file the cysteines are shown as reduced. These disulphide bonds are most likely artefacts that do not occur in physiological conditions between TBL1XR1 molecules. It is probable that these disulphide bonds helped with the crystallization by providing rigidity, and may be the reason why it was so difficult to reproduce these crystals.

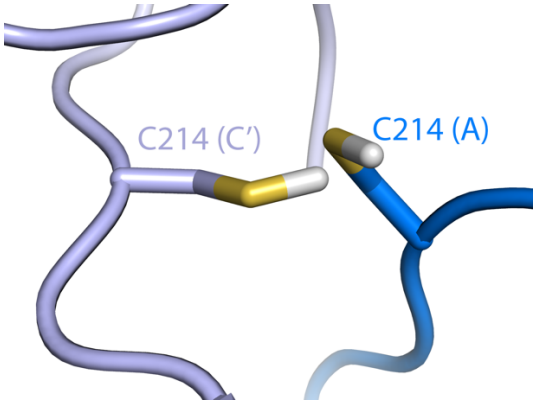

Figure 5.4.2. Disulphide bonds between adjacent molecules of TBL1XR1 in adjacent asymmetric unit cells. In this case, chain A of one unit cell makes a disulphide bond with a symmetry mate chain C

(marked C'). Because of issues with PDB file format, the cysteines are shown as reduced. Similar disulphide bridges occurred between B and D', C and A', and D and B'.

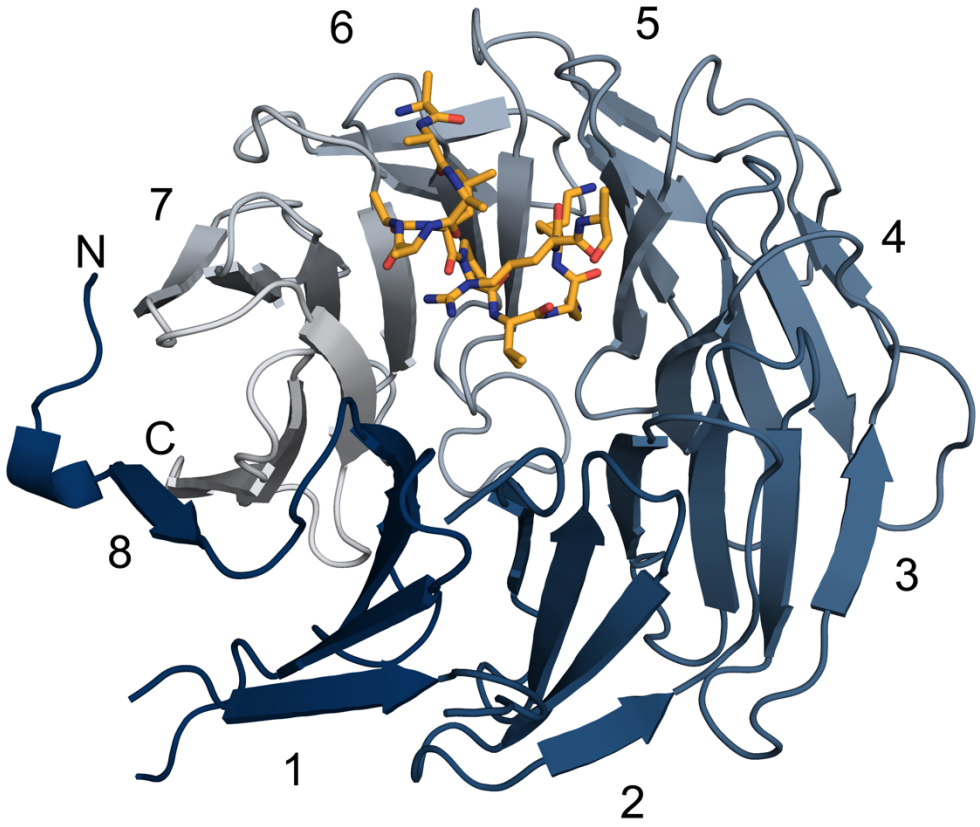

Figure 5.4.3. Cartoon representation of TBL1XR1 bound to MeCP2$_{285-309}$ peptide (orange). Each WD40 blade has been coloured a different shade of blue, starting with the darkest blue at the N-terminus and reaching pale blue/light grey at the C-terminus. The outermost beta strand belonging to blade 8 has been coloured blue to show its N-terminal location.

Like the human orthologue, mouse TBL1XR1 WD40 domain is an eight-bladed beta-propeller. The C-terminus of the domain packs against the N-terminus, closing the ring (Figure 5.4.3), which is a common feature of WD40 domains. The defining feature of WD40 repeats, however, is the DHSW tetrad, or a variation of thereof (Figure 5.4.4). This is a conserved motif in WD40 domains that forms an extensive hydrogen-bonding network and provides structural stability to the WD40 domain. Some of the hydrogen bonds in the DHSW tetrad have been found to be unusually strong, as one study found mutating a serine to an alanine

raised the folding free energy by around 9 kcal/mol, which would imply the contribution of each of the hydrogen bonds is around 4.5 kcal/mol (Wu et al., 2010).

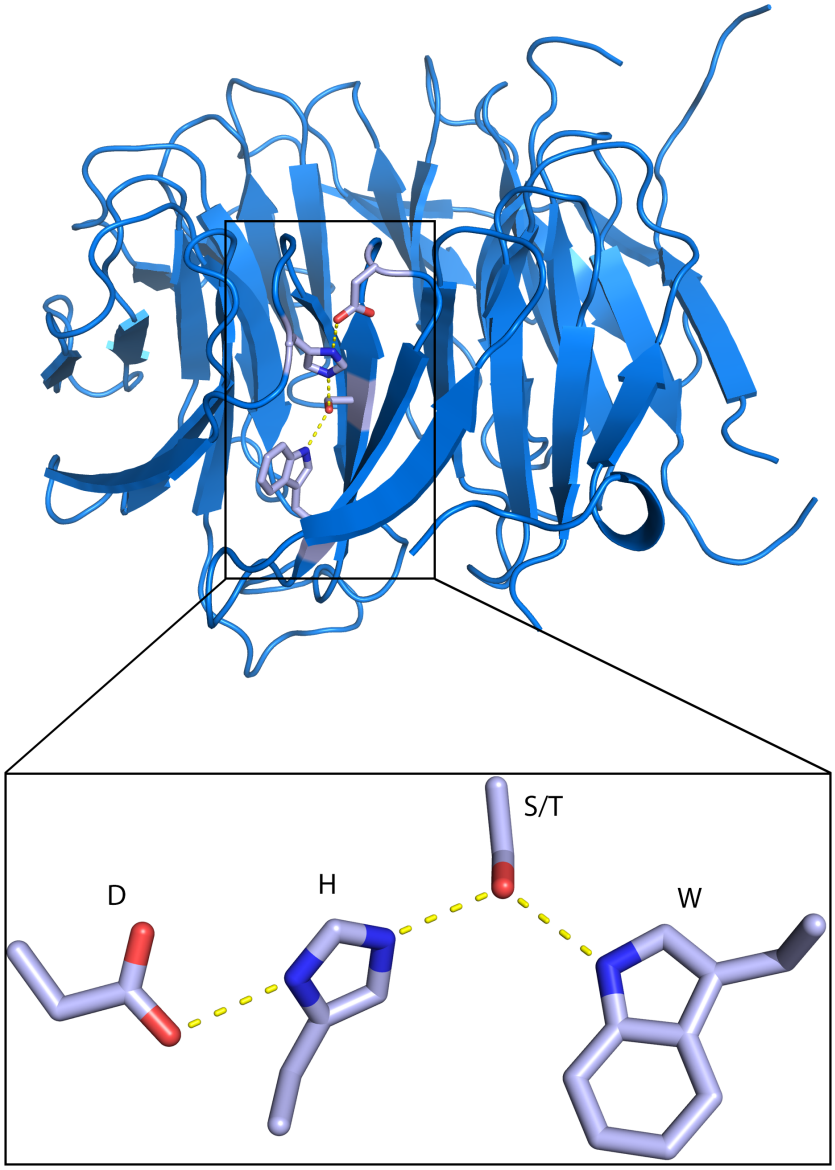

Figure 5.4.4. Example of a DHSW tetrad in a mouse TBL1XR1 WD40 domain. The name derives from a conserved arrangement of D-H-S/T-W residues that form intermolecular hydrogen bonds and provide stability to the WD40 fold.

There are six such DHSW motifs present in TBL1XR1. There is also a seventh (HSW) motif that is present in the human TBL1XR1, however it appears that the presence of the disulphide bonds in the mouse TBL1XR1 structure have caused the histidine side chain to shift, flipping out of its normal position (Figure 5.4.5). It is likely that this disruption would be detrimental to the stability of wild-type TBL1XR1, however the partial disruption of a single DHSW is

unlikely to lead to complete unfolding of the WD40 domain fold. It remains a possibility, though, that the disulphide bonds compensate for this.

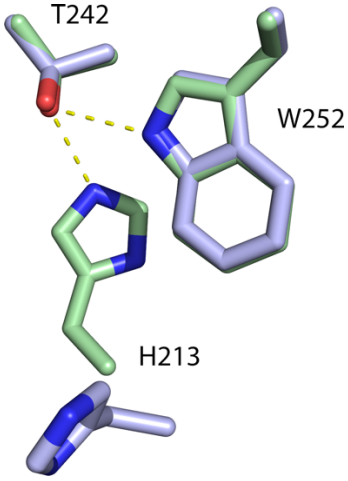

Figure 5.4.5. Overlay between human (green) and mouse (blue) seventh DHSW motif in TBL1XR1. There is no aspartate residue to complete the tetrad. In the mouse structure, His213 has been flipped out due to the presence of a disulphide bond between adjacent Cys214 and a Cys214 in a different asymmetric unit. Hydrogen bonds are shown between sidechains in the human TBL1XR1 structure.

The MeCP2 NID peptide is bound in the central channel on the top face of the WD40 domain. At the deepest point, the MeCP2 NID peptide is approximately 10 Å within this channel. As previously hypothesised, the electropositive MeCP2 NID peptide bound to a region of the WD40 domain that was highly electronegative (Figure 5.4.6). In the two copies of MeCP2, only residues 296-307 and 298-307 in chains E and F, respectively, were seen. Good electron density was present for the sidechains of residues 301-306 in the MeCP2 peptide. This indicates that the majority of the $MeCP2_{285-309}$ NID peptide is flexible and unlikely to be important in binding to TBL1XR1, and that the core region involved in binding spans residues 301-306.

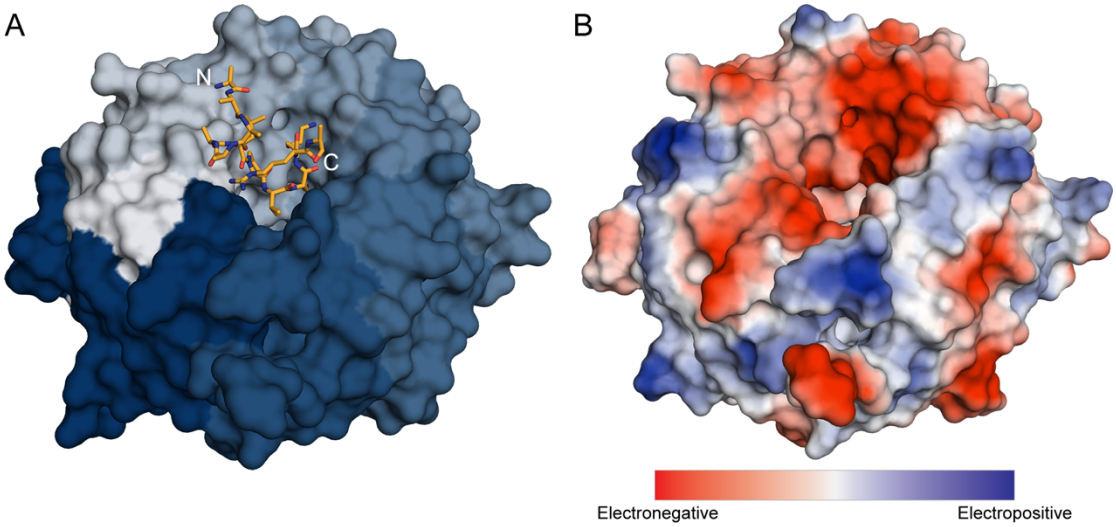

Figure 5.4.6. A) Surface representation of TBL1XR1-MeCP2 co-crystal. The MeCP2 peptide N and C-termini have been annotated. B) Electrostatic map from the same perspective.

No major rearrangements of the TBL1XR1 WD40 domain occur upon MeCP2 binding, exemplified by the r.m.s.d. of 0.34 Å between human (representing the unbound) and mouse TBL1XR1 (representing the bound) WD40 domain crystal structures. There were two amino acids, Y395^{TBL1XR1} and F420^{TBL1XR1}, which undergo conformational changes compared to the unbound form. These amino acids face outwards in the unbound structure, but rotate roughly 120˚ upon MeCP2 binding (Figure 5.4.7).

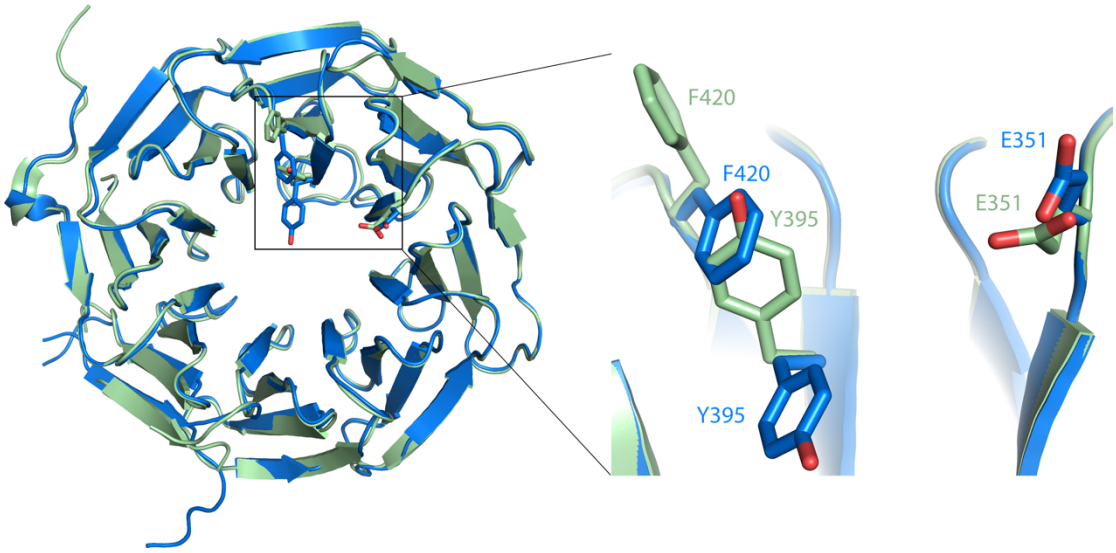

Figure 5.4.7. Left) Overlayed cartoon representations of human (green) and mouse (blue) TBL1XR1 WD40 domains. Right) Detailed view of the residues that move upon MeCP2 binding.

Together with Y446^{TBL1XR1}, they help create a hydrophobic pocket for the binding of P302^{MeCP2}. This will be discussed further in the next section. It is possible that the differences in those two amino acids have distinct conformations in human and mouse TBL1XR1, or that the

positions correspond to differences in the crystal packing. However, as these two amino acids are part of the MeCP2-TBL1XR1 interface, it is more likely that this is a real conformational change in response to MeCP2 binding. E351 also seems to move slightly upon MeCP2 binding, although the change in position is relatively small when compared to Y395 or F420. Nevertheless, the movement of E351 might represent an accommodation of the MeCP2 peptide. Unfortunately, it was impossible to compare the positions of the sidechains between bound and unbound mouse structures, as the two unbound WD40 domains in the crystal structure did not have electron density for these three sidechains. However, the r.m.s.d between the backbones of the bound and unbound mouse TBL1XR1 WD40 domains was 0.32 Å, which again shows that no major backbone rearrangements occur upon binding of MeCP2 NID peptide to the WD40 domain.

5.4.1 Interactions with TBL1XR1 are mainly mediated by four MeCP2 residues

The $MeCP2_{285-309}$ NID peptide makes extensive contacts with the TBL1XR1 WD40 domain. In particular, four of the MeCP2 residues found mutated in RTT (P302, K304, K305, and R306) appeared to be the major binding determinants. An overview of TBL1XR1 and MeCP2 residues involved in binding is shown in Figure 5.4.8.

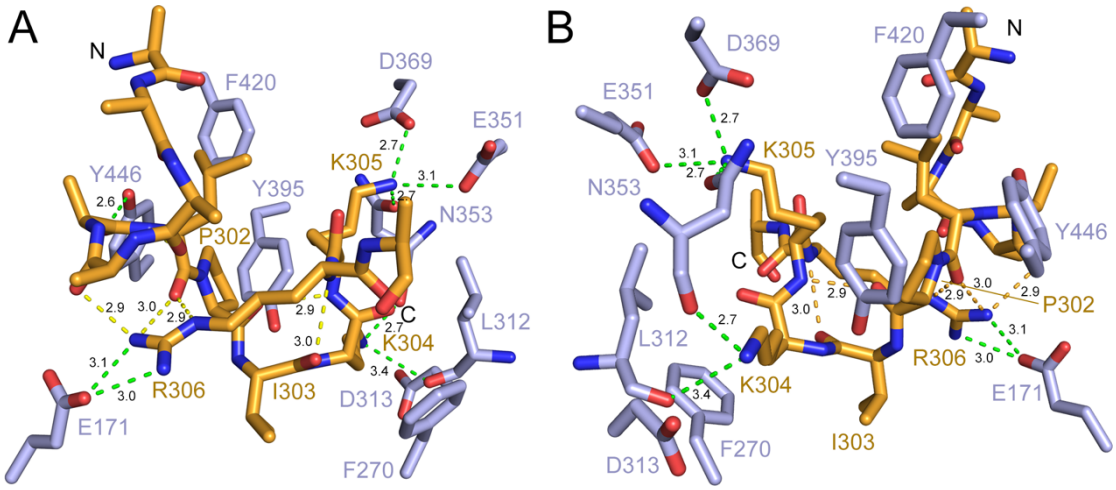

Figure 5.4.8. A) Key residues involved in MeCP2 (orange) and TBL1XR1 (light blue) interaction. Intramolecular hydrogen bonds are coloured orange while intermolecular hydrogen bonds are green. B) Same as A, but rotated 180 degrees around Y axis.

A bioinformatics tool (WDSP) has been developed to analyse the sequences of putative WD40 domain proteins and display both predicted DHSW tetrads and residues on the top face that are likely to be hotspots for binding (Wang et al., 2015, Wang et al., 2013, Wu et al., 2012).

Analysis of the mouse TBL1XR1 sequence predicted several residues that make contacts in the crystal structure (Table 5.4.1). While a few predicted residues were not involved in binding, the majority were. There were also three residues that were potentially involved in binding, however because the MeCP2 residues they contact do not have visible side-chains, it is difficult to assess their contribution. Nevertheless, the high level of agreement between the prediction and the actual crystal structure shows that this tool may be useful for analysing other WD40 domains.

Table 5.4.1. Residues in mouse TBL1XR1 WD40 domain predicted by WDSP to be important in binding proteins or peptides.

Residue	Makes contacts (Yes/No/Maybe)
E171	Yes
F173	Maybe
D227	No
Y245	Maybe
F270	Yes
L312	Yes
E351	Yes
E353	Yes
D369	Yes
E393	No
Y395	Yes
F420	Yes
Y446	Yes
F462	No
F487	Maybe

$P302^{MeCP2}$, which when mutated to arginine causes RTT, sits quite deep in the WD40 central channel and contacts $Y395^{TBL1XR1}$, $F420^{TBL1XR1}$, and $Y446^{TBL1XR1}$ (Figure 5.4.9). These three aromatic residues generate a hydrophobic pocket into which $P302^{MeCP2}$ can bind. $Y446^{TBL1XR1}$ also makes a hydrogen bond with backbone carbonyl of $V300^{TBL1XR1}$. Furthermore, $P302^{MeCP2}$ forms a hydrogen bond with backbone nitrogen of $K305^{MeCP2}$. In the case of RTT, it is conceivable that mutating a proline into an arginine not only causes major steric and electrostatic clashes with the hydrophobic pocket (Figure 5.4.9B), but also disrupts the

unique φ angle of proline. As the peptide binding is likely to be influenced by the tight backbone conformation provided by the proline, these two disruptions would most likely cause large conformational changes in the MeCP2 NID peptide and subsequent loss of binding to TBL1XR1.

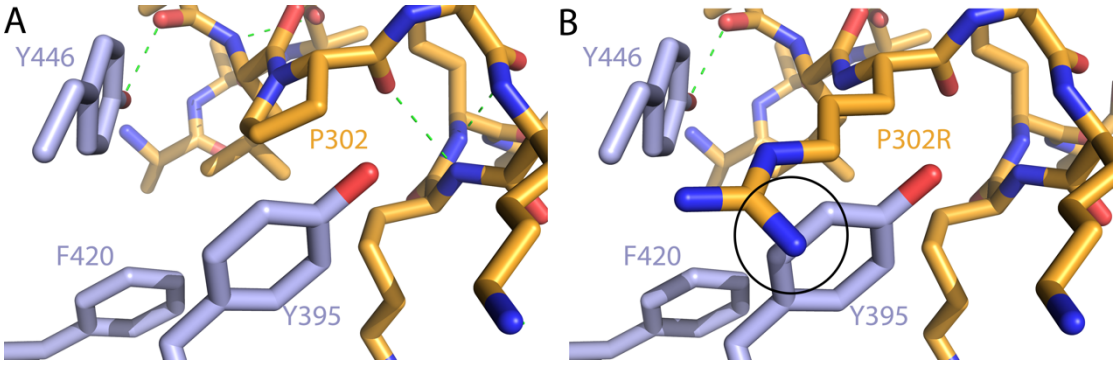

Figure 5.4.9. A) Binding site of P302^{MeCP2}. TBL1XR1 residues are light blue. B) The binding region with the P302R^{MeCP2} mutation is modelled. A clash between the P302R and Y395 is highlighted.

K304^{MeCP2} is involved in hydrogen bonding and salt bridges with the backbone carbonyl groups of N353^{TBL1XR1} and L312^{TBL1XR1} (Figure 5.4.10). While not directly involved in binding, it is likely that D313^{TBL1XR1} strengthens the binding of K304^{MeCP2} by enhancing an electronegative environment. Finally, the side-chain of F270^{TBL1XR1} is involved in van der Waals contacts with the side-chain of K304^{MeCP2}. When K304 is mutated to a glutamic acid in RTT, the loss of binding likely arises from electrostatic repulsion as the binding pocket of K304^{MeCP2} is highly electronegative.

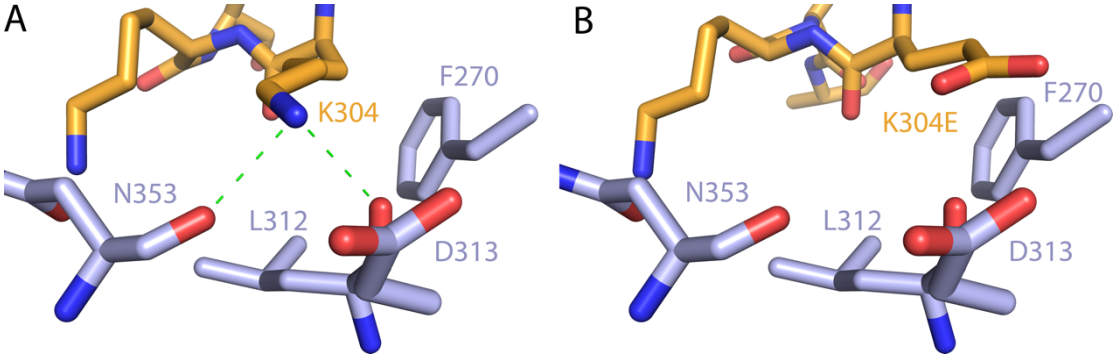

Figure 5.4.10. A) Binding site of K304^{MeCP2}. TBL1XR1 residues are light blue. B) The binding region with the K304E^{MeCP2} mutation is modelled.

K305^{MeCP2} forms hydrogen bonds with the sidechains of E351^{TBL1XR1}, N353^{TBL1XR1}, and D369^{TBL1XR1} (Figure 5.4.11). Like K304^{MeCP2}, the binding of K305^{MeCP2} to TBL1XR1 is probably a mix of hydrogen bonding and ionic interactions. The loss of binding in the case of RTT mutant

K305R^{MeCP2}, like P302R^{MeCP2}, is likely to cause a steric clash where the slightly larger arginine (and in particular the guanidium group) cannot fit into the binding pocket fit for a lysine.

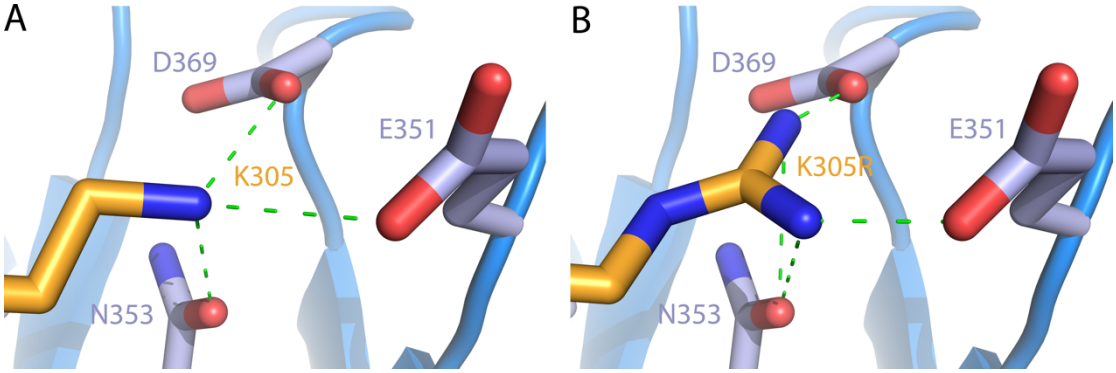

Figure 5.4.11. A) Binding site of K305^{MeCP2}. TBL1XR1 residues are light blue. B) The binding region with the K305R^{MeCP2} mutation is modelled.

Finally, R306^{MeCP2} makes extensive intra- and intermolecular hydrogen bonds with both E171^{TBL1XR1}, but also to the backbone carbonyls of T299^{MeCP2} and L301^{MeCP2} (Figure 5.4.12). These intramolecular hydrogen bonds are likely to be necessary for the peptide to maintain its shape, while the hydrogen bond to E171^{TBL1XR1} likely contributes to binding. When R306^{MeCP2} is mutated to cysteine in RTT (R306C^{MeCP2}), it is probably the loss of the stabilising hydrogen bonding network that causes a change in the MeCP2 NID peptide shape and the loss of binding to TBL1XR1, as the much shorter cysteine sidechain is both too far from the residues mentioned, and unable to form hydrogen bonds.

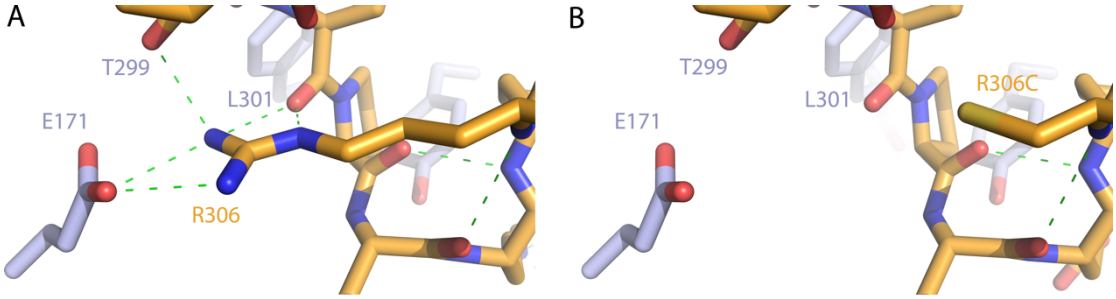

Figure 5.4.12. A) Binding site of R306^{MeCP2}. TBL1XR1 residues are light blue. B) The binding region with the R306C^{MeCP2} mutation is modelled.

It is unknown to what extent other residues in the MeCP2 NID are involved in the binding to TBL1XR1. Based on the structure it is reasonable to say that MeCP2 residues 301-306, and possibly V300, all contact TBL1XR1 to some degree, and that missense mutation of any of these amino acids would cause a loss or weakened binding of MeCP2 to TBL1XR1. Why there are no reports of other *de novo* mutations in the NID involved in RTT is unknown. It has been postulated that in the case of R306C, the arginine codons CGT and CGC can relatively easily

change to the TGT and TGC codons of cysteine through deamination of 5'-methylcytosine to thymidine, as in the context of CpGs the cytosine bases are often methylated. This could explain why R306C is over 16 times more common than P302R, the next most common *de novo* NID RTT mutation (RettBASE).

5.5 Purification of TBL1XR1 binding site mutants

In order to assess the importance of the TBL1XR1$_{134-514}$ residues that contact the MeCP2 NID peptide, a series of mutants were generated – E171A, E171Q, D313N, E351A, E351D, D369A, and Y446F. These were chosen based on the structure to study the effects of ionic interactions, hydrogen bonding, and van der Waals interactions. Because of the number of residues involved in binding, and their potential mutants, it was not feasible to mutate them all. Additionally, C214S was made for the purposes of crystallisation, in the hopes that by abolishing the formation of disulphide bridges between adjacent molecules of TBL1XR1, it would be easier to develop a protocol for reproducible crystals. Both the site directed mutagenesis and subsequent protein preparations were carried out by Ceitidh Taylor, an undergraduate student under my supervision from January-April 2016. The mutant proteins were expressed in Sf9 cells, just like the wild type TBL1XR1$_{134-514}$. Initial pulldown assays showed that all mutants expressed to comparable levels to the wild-type TBL1XR1 (Figure 5.5.1)

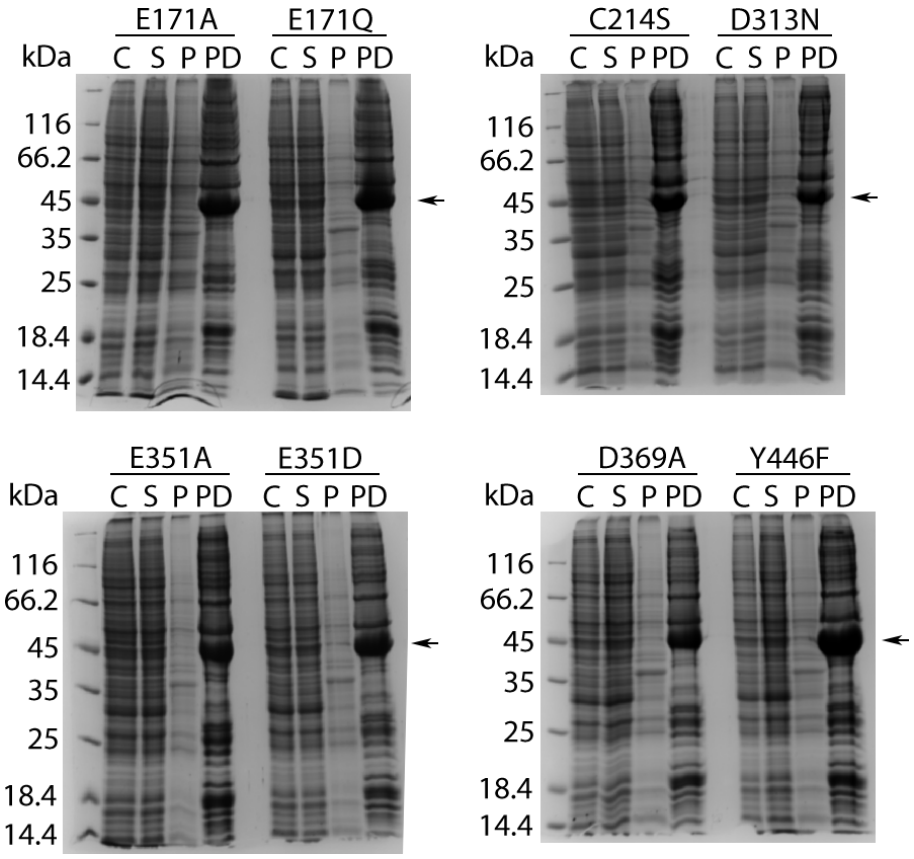

Figure 5.5.1. Pulldown assays of all TBL1XR1$_{134-514}$ mutants. Black arrows mark the expected size of the TBL1XR1 mutants.

As there was no reason to suspect these mutations would require major alterations to the purification protocol, the purification of mutants was carried out using the same protocol as wild-type TBL1XR1. Representative SDS-PAGE gel from the SEC step together with the SEC chromatogram from E171A purification are shown in Figure 5.5.2 and Figure 5.5.3. Indeed, all of the mutants could be purified following the wild-type TBL1XR1 protocol (Chapter 3.8) and all mutants produced similar yields and purity when compared to wild-type (4-5 mg/20g of wet cell pellet).

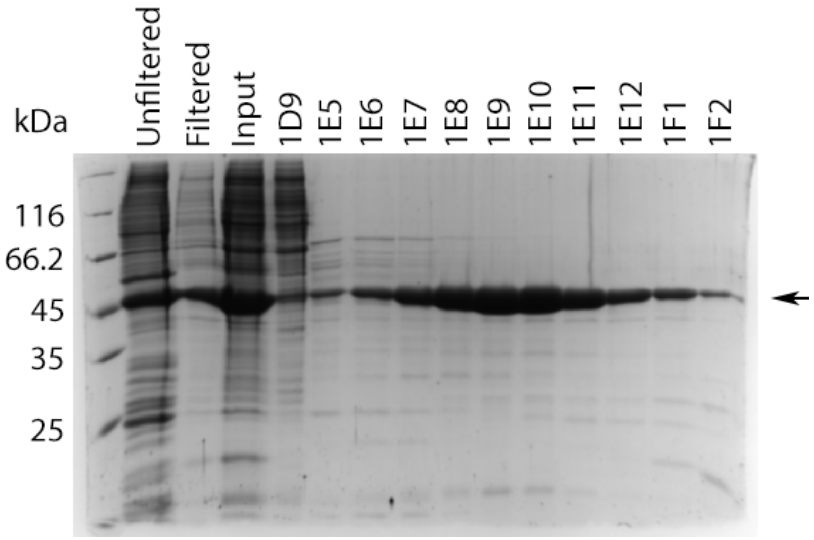

Figure 5.5.2. SDS-PAGE gel of SEC step of TBL1XR1$_{134-514}$ E171A purification. TBL1XR1$_{134-514}$ E171A has been marked with a black arrow.

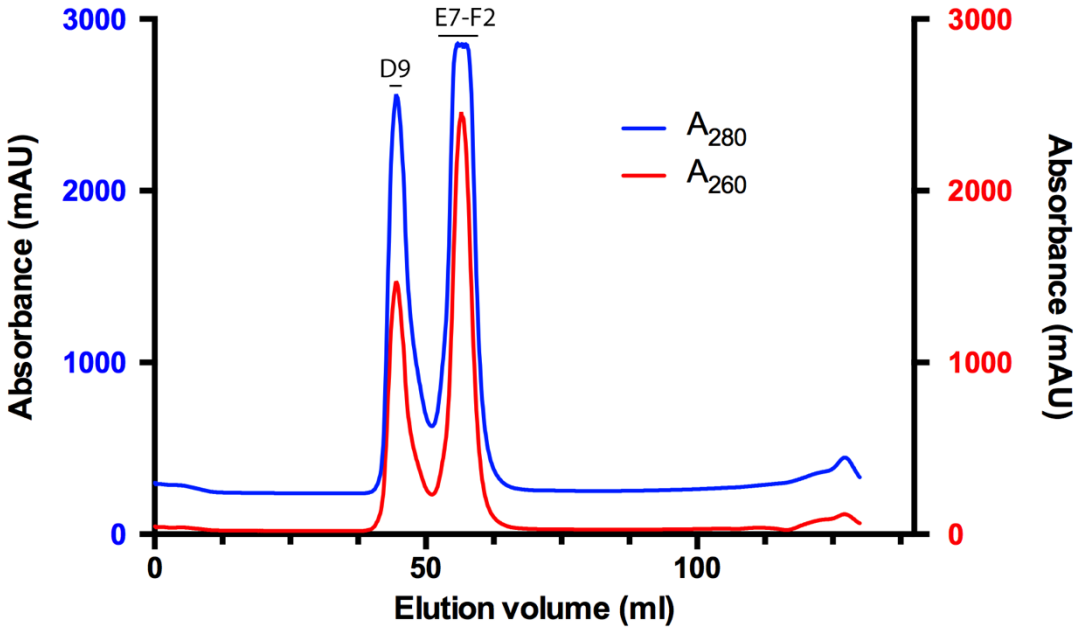

Figure 5.5.3. SEC chromatogram of TBL1XR1$_{134-514}$ E171A purification.

5.6 Characterisation of TBL1XR1 binding site mutants

To assess whether the mutations introduced into TBL1XR1 caused gross destabilisation of the WD40 fold, melting temperatures of each of the mutants were measured using TDA and compared to the wild-type TBL1XR1 (Figure 5.6.1). No mutant had a change in the melting temperature of more than 3°C. As the wild-type protein has a relatively high melting

temperature (66°C), a three degree change to 63˚C is relatively small and the protein is likely to still be stable. Therefore, none of these point mutations caused major destabilisation of the protein. It was therefore possible to use the mutants to study the effects of the point mutations on the binding of MeCP2 to TBL1XR1.

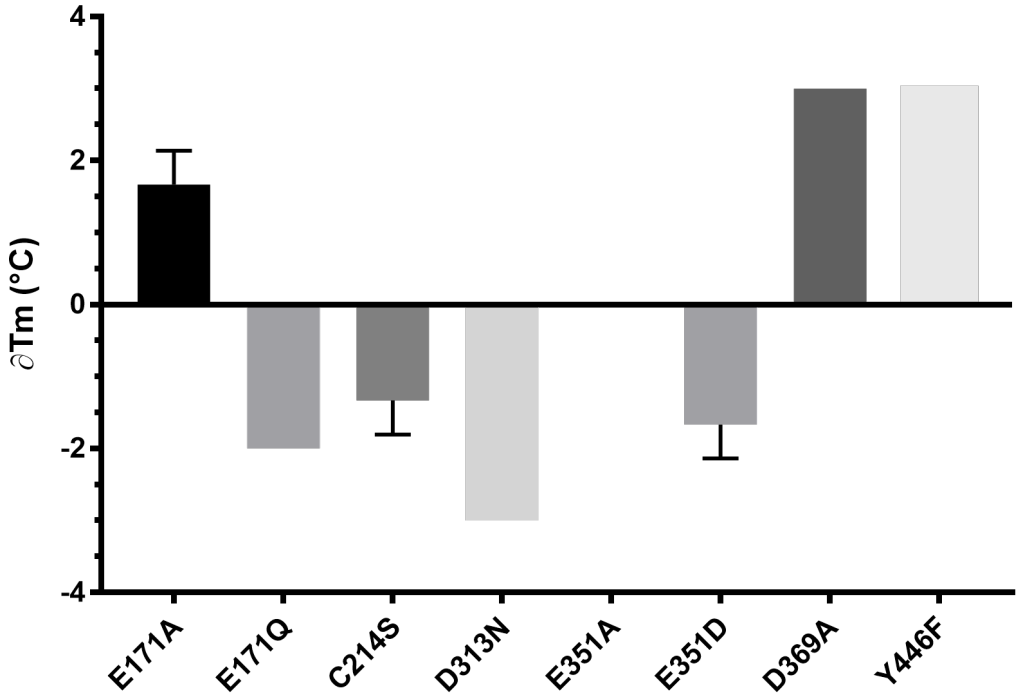

Figure 5.6.1. Change in melting temperatures of TBL1XR1$_{134-514}$ mutants when compared to wild-type TBL1XR1 (set as reference, 66°C). Error bars are one standard deviation.

Pulldowns using insect cell lysates and biotinylated MeCP2$_{285-313}$ peptides were performed to see whether any of these mutants abolished the binding of MeCP2 to TBL1XR1. TBL1XR1$_{134-514}$ wild-type protein was used as a positive control, and MeCP2$_{285-313}$ R306C and K305R peptides were used as negative controls. Wild-type TBL1XR1, as well as C214S, E351D, and Y446F mutants all bound to wild-type MeCP2 peptide, whereas the other mutants failed to do so (Figure 5.6.2). None of these TBL1XR1 constructs bound to R306C and K305R peptides (Figure 5.6.3 and Figure 5.6.4), although previously it was hypothesised that K305R^{MeCP2} might bind to E351D^{TBL1XR1} as the shorter side-chain of the aspartic acid might be able to accommodate the larger side-chain of the arginine.

To confirm that the protein that was co-precipitated with wild-type MeCP2 peptide was indeed TBL1XR1, MALDI-TOF-TOF analysis was performed. This revealed that both of the bands pulled down by MeCP2 peptide were TBL1XR1, and both were annotated as mouse TBL1XR1 with scores of 146 and 140, and coverage of 43% and 49% for the upper and lower

bands, respectively. The presence of two bands could occur if some of the N-terminal regions of the recombinant mouse TBL1XR1 protein were proteolysed and the lower band seen corresponds to a stable fragment, such as the intact WD40 domain. Alternatively, the lower band may be an endogenous *S. frugiperda* TBL1XR1. As the genome of *S. frugiperda* has not been sequenced, this could not be confirmed by sequence comparison. Comparing TBL1XR1 to the sequence of another invertebrate TBL1X, a *Drosophila melanogaster* protein EBI, showed the two proteins are approximately 80% conserved in sequence in the WD40 domains. However, the discrepancies were interspersed in sequence. As the peptides identified in mass-spectrometry were on average more than ten amino acids long, it is unlikely that the peptides would be identical to the sequence of *S. frugiperda,* and therefore the second band probably belonged to the recombinant mouse TBL1XR1. As it did not impede the interpretation of the results or further experiments, it was not investigated in detail.

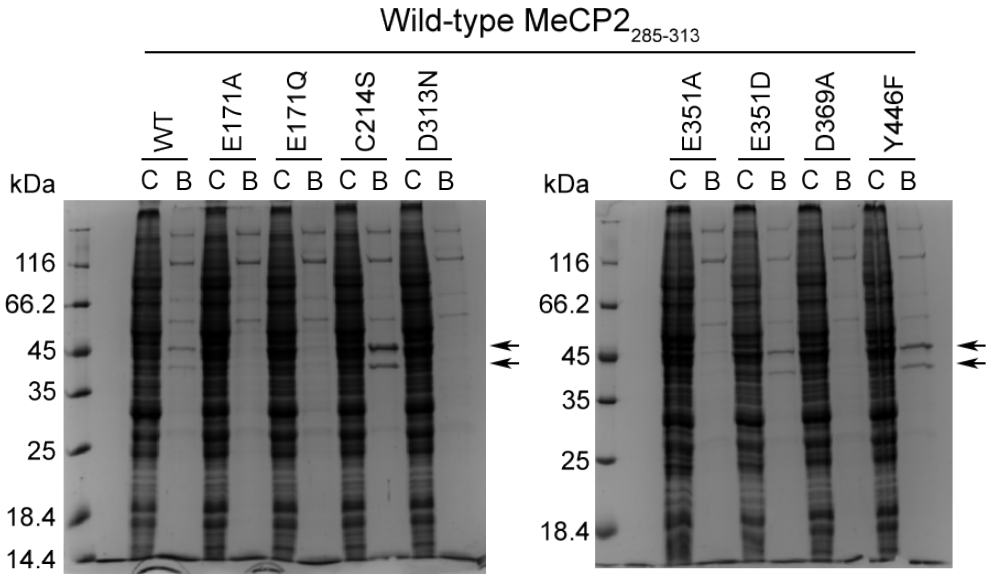

Figure 5.6.2. Pulldown assays of wild-type and mutant TBL1XR1$_{134-514}$ from crude cell lysates using N-biotinylated wild-type MeCP2$_{285-313}$. TBL1XR1$_{134-514}$ constructs that were precipitated from the crude lysate have been marked with a black arrow.

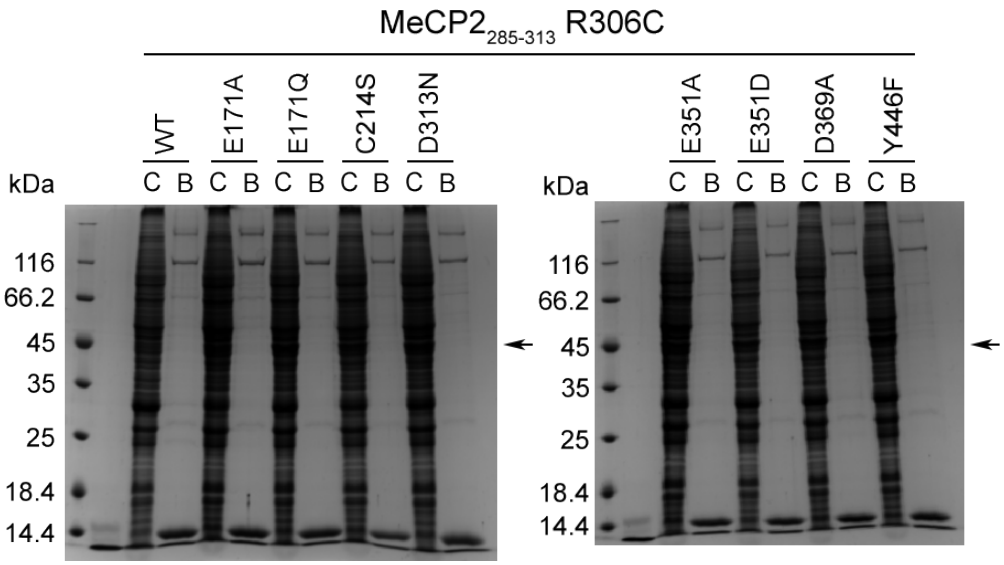

Figure 5.6.3. Pulldowns of wild-type and mutant TBL1XR1$_{134-514}$ from crude cell lysates using N-biotinylated MeCP2$_{285-313}$ R306C. Black arrows indicate the location of the expected TBL1XR1 band.

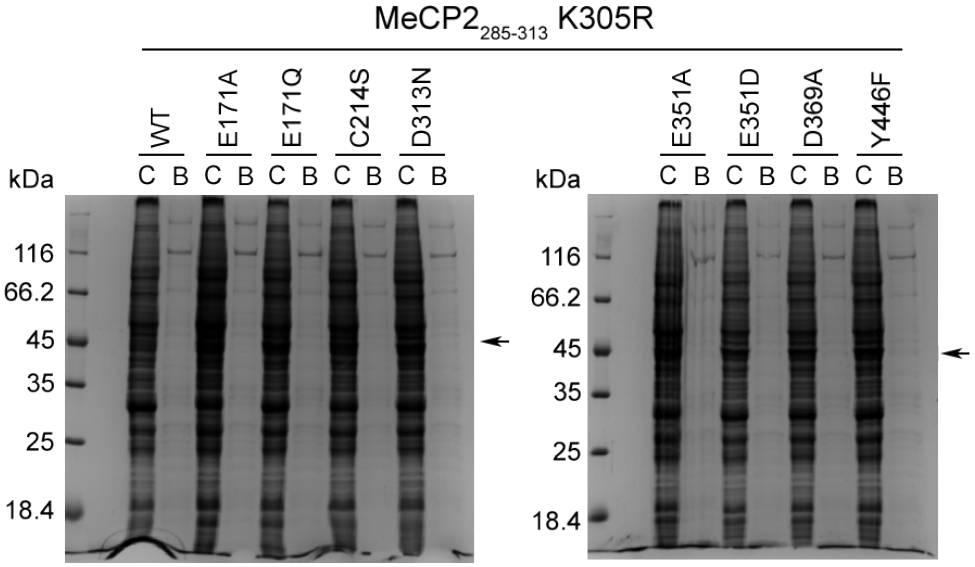

Figure 5.6.4. Pulldowns of wild-type and mutant TBL1XR1$_{134-514}$ from crude cell lysates using N-biotinylated MeCP2$_{285-313}$ K305R. Black arrows indicate the location of the expected TBL1XR1 band.

To assess to what extent these mutations in TBL1XR1 affect the binding to MeCP2 NID, SPR assays were performed with MeCP2$_{285-313}$, MeCP2$_{285-309}$ and MeCP2$_{285-313}$ R306C peptides, and the K_D values calculated. The measured RUs have been plotted and shown in Figure 5.6.5 and Figure 5.6.6 (raw data is in appendix 10.1). All of the mutants except for D313N gave a measurable response for MeCP2$_{285-313}$. When using MeCP2$_{285-309}$ peptide, a measurable response could be obtained for each of the mutants, except D313N.

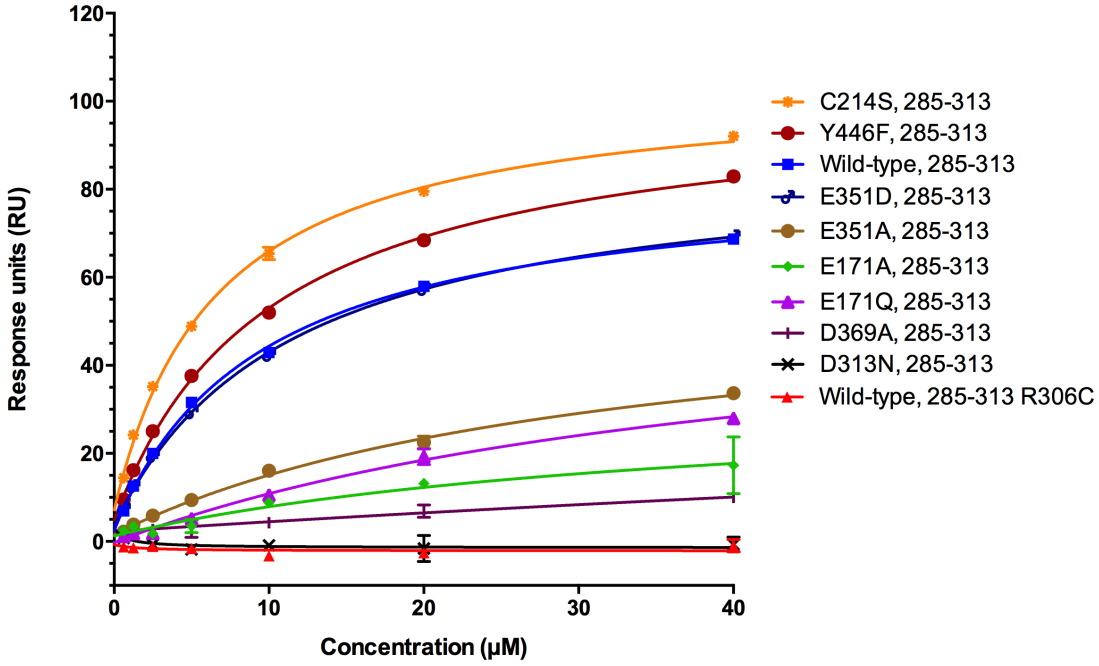

Figure 5.6.5. Binding curves for all of the TBL1XR1 mutants binding to MeCP2$_{285-313}$. MeCP2$_{285-313}$ R306C has been included as a negative control. Error bars are one standard deviation. Data has been fit to Langmuir isotherm using non-linear regression.

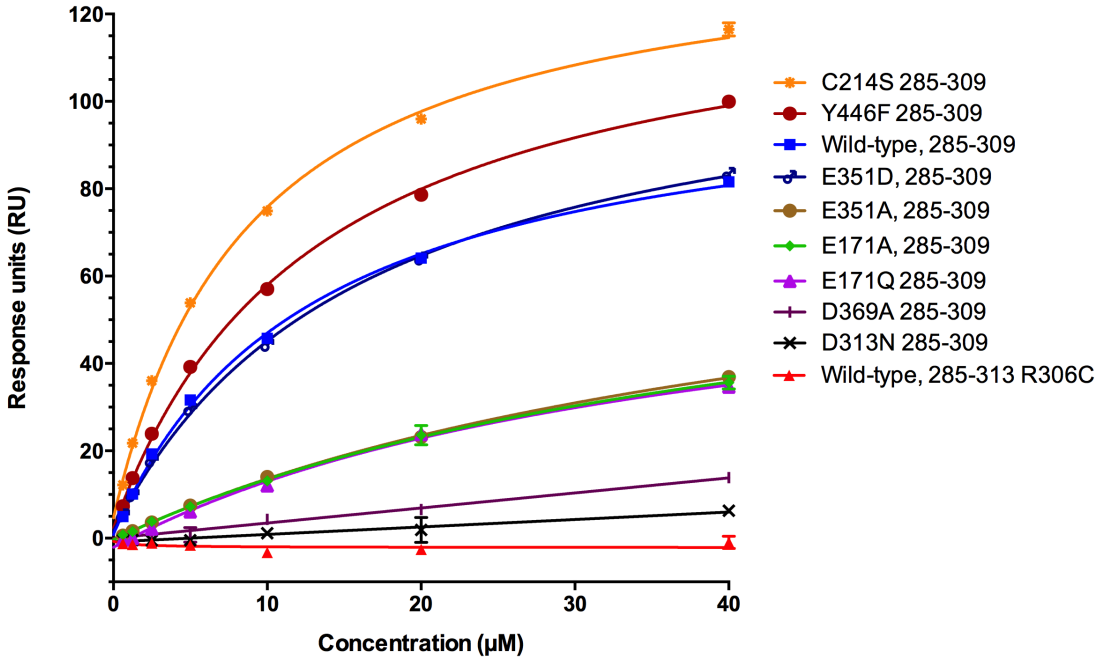

Figure 5.6.6. Binding curves for all of the TBL1XR1 mutants binding to MeCP2$_{285-309}$. MeCP2$_{285-313}$ R306C has been included as a negative control. Error bars are one standard deviation. Data has been fitted to a Langmuir isotherm using non-linear regression.

The data in these two figures were fit to Langmuir isotherm using non-linear regression, which gave the K_D values for these mutants (Table 5.6.1). E351D and Y446F bound with similar affinity as wild-type TBL1XR1. As expected C214S, which is not near the MeCP2 binding site, bound with wild-type affinity too. E171A and E171Q, as well as E351A showed much weaker interactions with either of the MeCP2 peptides, while D313N and D369A produced responses that were so low that K_D could not be calculated. This is because the value of maximum response unit (which is necessary for K_D calculation) was negative, or because the small change in response upon binding produced K_D values in the range of 10^{16} M. From these data several things can be concluded. First, as both E171A and E171Q showed hugely diminished binding to MeCP2, the salt bridge formed between $E171^{TBL1XR1}$ and $R306^{MeCP2}$ seems to be more important than the hydrogen bonding between these two residues. Secondly, E351A had a much lower affinity compared with E351D. Because E351Q mutant was not generated, it is not possible to assess the importance of the ionic interactions over hydrogen bonding between $E351^{TBL1XR1}$ and $K305^{MeCP2}$. However, it is probable that E351 contributes to the local electronegative environment of $K305^{MeCP2}$ binding pocket in similar fashion to E171, in addition to hydrogen bonding. Finally, both D313N and D369A failed to bind to either MeCP2 peptides. The result of D313N was surprising as it was not expected that this mutation would abolish the interaction completely. Rather the expected effect was something similar to E171 mutants. As mentioned previously, D313 is not directly involved in binding to MeCP2, but like E171A and E351, it forms a part of a continuous electronegative surface. It is therefore possible that D313N grossly disturbs the local electrostatic environment of TBL1XR1, preventing $K304^{MeCP2}$ from binding. Alternatively, D313N might help position L312 and N353 so that they are able to contact MeCP2. In the case of D369A, it is again most likely that a combination of loss of hydrogen bonds and ionic interactions contributed to the observed loss of binding. Finally, mutating Y446 into a phenylalanine had negligible effect on the K_D values, indicating that the hydrogen bond made with the MeCP2 backbone contributes little to binding, and that the aromatic side-chain that contacts $P302^{MeCP2}$ likely has a larger contribution, although this was not confirmed with a Y446A mutation.

Table 5.6.1. Calculated K_D values for all the mutants from SPR data. Wild-type has been included as a reference. N/A means the numbers calculated were not meaningful and represent non-binders; SEM is standard error of the mean.

	$MeCP2_{285-313}$	**$MeCP2_{285-309}$**
	K_D ± SEM (μM)	**K_D ± SEM (μM)**
Wild-type	9.5 ± 0.5	12.9 ± 0.8
E171A	39.0 ± 30.7	45.8 ± 7.1
E171Q	44.3 ± 10.1	37.9 ± 4.9
C214S	6.6 ± 0.4	8.9 ± 0.5
D313N	N/A	N/A
E351A	32.4 ± 4.9	49.6 ± 4.3
E351D	11.4 ± 0.5	16.3 ± 0.7
D369A	N/A	N/A
Y446F	10.1 ± 0.5	13.0 ± 0.7

5.7 Discussion

During this part of the project TBL1XR1-MeCP2 was successfully co-crystallised after optimising the crystallisation conditions based on leads from sparse-matrix screening. The obtained crystals diffracted to a resolution of 2.5 Å. There were four molecules of TBL1XR1 in the asymmetric unit, and two of them had MeCP2 NID peptide bound. Previously only the structure of MeCP2 MBD with and without DNA had been solved. This chapter presents the first reported structure of MeCP2 NID.

The mouse TBL1XR1 WD40 domain is an eight-bladed beta-propeller that is structurally similar to the previously published human TBL1XR1 WD40 domain. The MeCP2 NID peptide was bound in the central channel on the top face of WD40 domain, in line with previous biochemical data that showed that a residue on the top face (E351 in TBL1XR1, E364 in TBL1X) is required for the binding of MeCP2. Analysis of the bound peptide revealed that the four amino acids that are found mutated in *de novo* cases of RTT (P302, K304, K305, and R306) made extensive contacts with the TBL1XR1 WD40 domain. These RTT mutations likely abolish the binding of MeCP2 NID to TBL1XR1 through a combination of steric clashes (P302R and

K305R), electrostatic repulsion (K304E), and loss of unique peptide shape through abolishing intramolecular hydrogen bonds (R306C).

To assess the importance of TBL1XR1 residues on the binding of the MeCP2 peptide, several of these residues were mutated and the affinities of these mutant proteins towards MeCP2 measured using surface plasmon resonance assays. The effects of these mutations were also confirmed with pulldown assays using cell lysates. $TBL1XR1_{134-514}$ C214S, E351D, and Y446F mutations all had similar binding affinities to wild type. E171A, E171Q and E351A all displayed greatly weakened binding, and both D313N and D369A failed to bind completely. Overall this indicated that the electronegative surface is largely responsible for the binding of MeCP2 to TBL1XR1, with hydrogen bonds contributing less to binding.

The crystal structure provides an atomic level understanding of how MeCP2 NID binds to TBL1XR1 WD40 domain. Through this, it was possible to explain why known RTT mutations in the NID abolish this interaction. It was also demonstrated that MeCP2 binding could be abolished by reciprocal mutations in TBL1XR1. This raised a question whether mutations in TBL1X or TBL1XR1 could cause RTT. While mutations in TBL1XR1 have been reported to cause West syndrome (G70D) (Saitsu et al., 2014), Pierpont syndrome (Y446C) (Heinen et al., 2016), and uncharacterised autism-spectrum disorder (L282P) (O'Roak et al., 2012), none of these have been classified as RTT. The Deciphering Developmental Disorders project (Deciphering Developmental Disorder), which aims to identify mutations that cause developmental delay, has identified several *de novo* missense mutations in TBL1XR1 listed in Table 5.7.1. There is also a frameshift mutation that causes the protein to be truncated at amino acid 275. Interestingly, to date no single nucleotide variations have been found in TBL1X that lead to disease.

Table 5.7.1. *De novo* missense mutations found in TBL1XR1 through Deciphering Developmental Disorders project.

Mutation	Phenotype	Sex
H213Q	Global developmental delay	Male
A311P	Nervous system abnormality	Male
D328G	Global developmental delay and epilepsy, although also has mutation in	Male

	SCN2A gene which encodes for a Na^{2+} channel	
D369E	Severe intellectual disability	Male
D370Y	Intellectual disability	Female
H441R	Global developmental delay	Male
P444R	Moderate global developmental delay	Female

Mutations H213Q, D328G, D370Y, and H441R are all part of the integral DHSW tetrads described earlier, and it is likely that the effects seen are due to both disruption of the hydrogen bonding in the DHSW tetrad, as well as changes in local geometry. Whether these mutations also cause gross destabilization of the WD40 fold is unknown.

Ex vivo assays performed by Dr. Matthew Lyst have shown that both D369E and P444R mutations abolish the binding of MeCP2 to TBLR1 (personal communication). P444 is found in a loop between two beta strands near MeCP2 binding site. The P444R mutation likely disrupts both the loop conformation, as well as MeCP2 binding through steric clashes with the MeCP2 peptide. Similarly, D369 is directly involved in binding to MeCP2 peptide and mutation of an aspartate to a longer glutamate (D369E) likely clashes with $K304^{MeCP2}$. While the Y446F mutation did not abolish the binding of MeCP2 to TBL1XR1, and Y446A mutation was not tested, it is possible that Y446C mutation does prevent binding as it is unable to contribute to the hydrophobic binding pocket of $P302^{MeCP2}$.

L282 and A311 are both found in beta strands. The mutation of either of those residues into a proline is very likely to disrupt the secondary structure and cause destabilisation of the WD40 blades. Finally, G70D mutation exists in the region between LisH and WD40 domain. This region has been shown to interact with NCoR/SMRT, and it is possible that this mutation blocks this interaction. However, it is also possible that there are other interacting protein partners of TBL1XR1 and that these mutations interfere with yet unknown interactions.

If at least some of these mutations indeed prevent interaction of TBL1XR1 with MeCP2, the question is why do these mutations not cause RTT, but a related phenotype. This can be explained in several ways. As TBL1XR1 is autosomal but MeCP2 is not, mutations in TBL1XR1 could be compensated by the paralogue TBL1X. However, specific tissue expression data for TBL1X and TBL1XR1 is not available and it is not known to what extent these two proteins are expressed in the same tissues. In addition, it remains unknown whether these proteins

have unique roles or not. The different phenotypes may also arise from the fact that MeCP2 is subject to X-inactivation and therefore the female patients are mosaic with some cells containing wild-type MeCP2 and some cells containing mutated MeCP2. Because TBL1XR1 is autosomal, such mechanism does not occur and mutations in the gene may instead cause haploinsufficiency in all cells. While this question will not be pursued in the scope of this project, hopefully these questions can be answered in the future.

6 Characterisation of MeCP2 'minigene' protein

6.1 Introduction

While the sequence of MeCP2 is widely conserved throughout vertebrates, the purpose of the regions beyond the MBD and NID are still debated with conflicting views between different research groups as to what these regions do. As a majority of the RTT mutations occur within the MBD or the NID with only 3 known *de novo* RTT mutations outside these areas, the question was whether these regions are involved in RTT pathogenesis. To assess this, a number of constructs were made by Rebekah Tillotson from Adrian Bird's lab at the Wellcome Trust for Centre of Cell Biology, University of Edinburgh. These MeCP2 constructs had some or all parts of the protein, other than the MBD and NID, deleted (Figure 6.1.1).

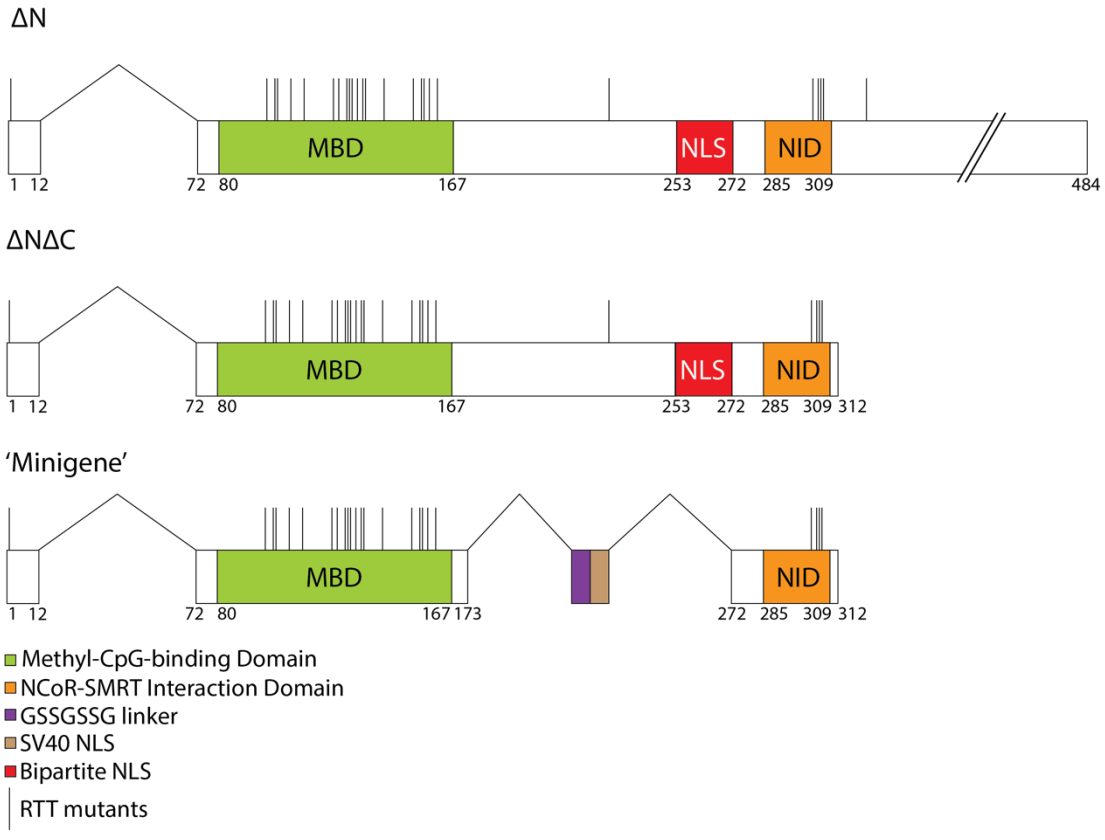

Figure 6.1.1. Domain organisation of the three MeCP2 constructs. All the constructs that were introduced into mice also contained a C-terminal EGFP tag not pictured here.

She generated transgenic mice containing these constructs and observed their phenotypes. Of particular interest was the construct dubbed the 'minigene'. This was the most minimal

construct made, containing only the MBD, NID, and a nuclear localisation signal (NLS) from SV40. It also had a C-terminal EGFP tag not shown here. While the knock-in mice with the constructs that did not have either the N-terminus (dubbed ΔN), or the N- and C-termini (dubbed ΔNΔC) were phenotypically similar to wild type mice, the minigene mice had a mild RTT phenotype. A 'bridge' model proposed by Lyst et al, 2013 hypothesised that MeCP2 acts as a molecular bridge between methylated DNA and NCoR/SMRT, and that the loss of either of the interactions causes RTT by abolishing the bridging function. Based on this model, the mice bearing the minigene protein could be displaying mild RTT symptoms due to the disruption of the bridging function, which could happen for several reasons. First, the deletion of the conserved linker region in the MeCP2 protein could have disrupted the functions of MBD and NID by either abolishing or changing their binding affinities to their respective partners. The close proximity of the MBD to the NID in the minigene protein could have had a similar effect on these interactions. It was also possible that the deletion of the conserved linker region between MBD and NID was responsible for the phenotype by disrupting a reported AT-hook (amino acids 265–272). Finally, further research performed by Rebekah Tillotson showed that the minigene may not be as stable as wild type protein *in vivo*, as the protein levels were reduced by approximately 50% when compared to wild-type. However, the mRNA levels were approximately two times higher than that of wild-type MeCP2, which could mean there are issues with translation. In either case the minigene protein could be functionally similar to wild-type, but the decreased levels of the protein in cells could lead to the mild phenotype seen.

In this chapter the functionality of MBD and NID in the context of the minigene were examined. Electrophoretic mobility shift assays (EMSAs) were conducted to see whether the MeCP2 minigene protein is able to bind to both DNA and TBL1XR1. In addition, a hypothesis that the MeCP2 is just a molecular bridge between DNA and TBL1XR1 was tested by examining the affinity of MeCP2 towards DNA in the absence and presence of TBL1XR1 using fluorescence anisotropy assays. It has been shown that MeCP2 binds cooperatively to DNA (Ghosh et al., 2008). DNA-binding proteins can be allosterically regulated through changes in their affinity for interaction partners upon DNA binding (Lefstin and Yamamoto, 1998). It is possible that MeCP2 displays similar effect by changing the affinity of the NID to TBL1X/TBL1XR1 when bound to DNA, the affinity of the MBD to DNA when bound to TBL1X/TBL1XR1, or both. This would further the understanding about the role of MeCP2 in RTT.

6.2 Construct design and cloning

As mentioned previously, the initial MeCP2 minigene construct was cloned by Rebekah Tillotson for the purposes of generating a knock-in mouse that only contained the MBD and NID of MeCP2, as well an NLS to make sure the protein can be transported into the nucleus. Since there are reports indicating that the endogenous MeCP2 NLS is an AT-hook (Baker et al., 2013), it was replaced by an SV40 NLS to abolish DNA binding provided by this AT hook. The recombinant MeCP2 minigene construct with an N-terminal 6x His tag followed by a 3C cleavage site was cloned into a bacterial expression vector. For the purposes of tracking the protein in cells, the knock-in gene also contained a C-terminal EGFP tag, however in the bacterial expression vector the EGFP tag was removed as it was not necessary for the biochemical assays.

6.3 Expression tests in Escherichia coli

The MeCP2 minigene protein was expressed overnight at 18˚C in BL21(*DE3*)-derived cells. The cell lysates were used in pulldown assays to test for protein expression and while there was a band that corresponded to the size of MeCP2 minigene protein (Figure 6.3.1), the cells themselves looked unhealthy and the biomass from 2 litres of cultures was low (around 5 g of wet cell pellet per 2 L). As the minigene contains the MBD, the minigene most likely binds non-specifically to the bacterial genome and interferes with normal cell function, causing poor cell health. Therefore the protocol developed for expressing MeCP2 methyl-CpG-binding domain (MBD) was used to express the minigene instead (Klose and Bird, 2004).

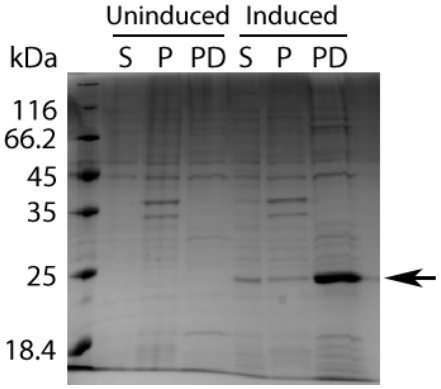

Figure 6.3.1. Small-scale expression test of MeCP2 minigene that was grown overnight at 18˚C. A band corresponding to the molecular weight of the minigene was present in the pulldown fraction of the induced sample (black arrow). S – soluble, P – pellet, PD – pulldown.

Using the method by Klose *et al.*, cells grown at 30 ˚C were monitored throughout the 3 hours of induction period. After one to two hours post-induction, the cells had entered growth arrest as evidenced by the lack of increasing OD_{600} value. Further samples were taken every hour until the cells were harvested. Pulldowns were performed on all the induced samples and while the cells stopped doubling, the levels of MeCP2 minigene increased. Most importantly, even though the biomass was still on the low side (around 10 g of wet cell pellet per 2 L of culture), it was twice as much as for the overnight growth, produced comparable amounts of protein in pulldown assays, and was much quicker. It was therefore decided to use this expression protocol for MeCP2 minigene protein.

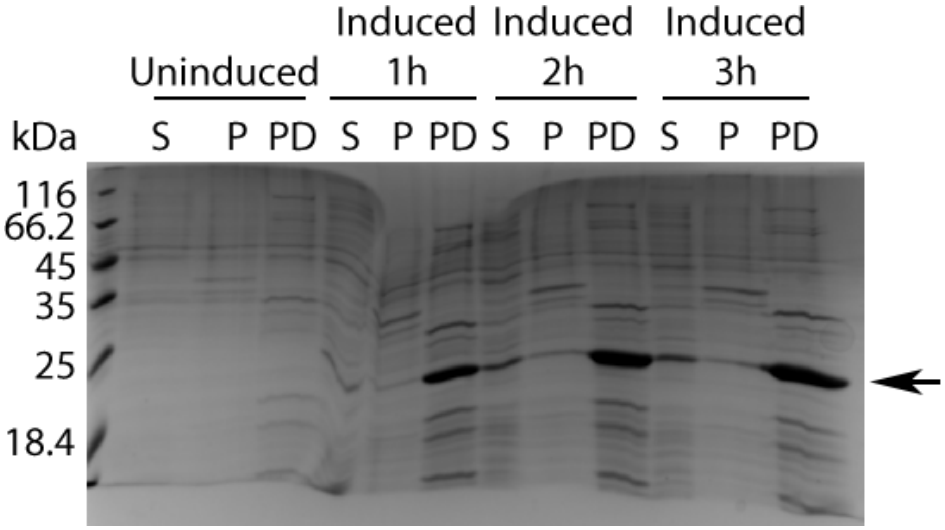

Figure 6.3.2. Small-scale expression test of MeCP2 'minigene' that was grown for 3h at 30˚C. S – soluble, P – pellet, PD – pulldown.

6.4 Protein purification

Analysing the protein sequence with ProtParam indicated that the predicted pI of the protein was 10.2. Consequently, the initial proposed protocol included IMAC as the first step, followed by cation-exchange chromatography, and then finally size-exclusion chromatography to further increase the purity of the protein as well as to verify that it exists as a single species.

Following cell lysis and clearing of the supernatant through centrifugation, the lysate was added to Ni^{2+}-NTA resin to bind the MeCP2 minigene protein. Because it was not known how well the minigene protein binds to the resin, five elutions at increasing imidazole concentrations were carried out. The MeCP2 minigene protein eluted off the resin in all the conditions, so in subsequent experiments a step elution with 500 mM imidazole was used.

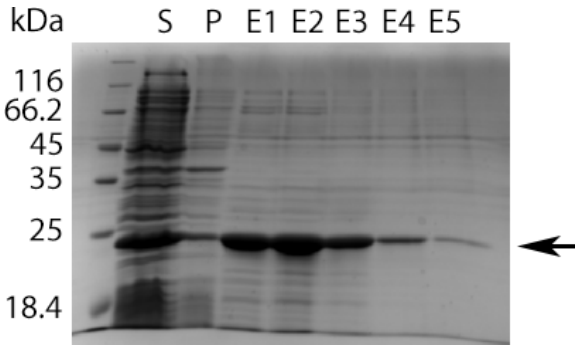

Figure 6.4.1. Batch pulldown of MeCP2 'minigene' protein using Ni^{2+}NTA resin. The protein of interest is marked with black arrow. Elutions were done by increasing the concentration of imidazole by 100 mM each time, starting from 100 mM in E1, and ending with 500 mM in E5. S – Soluble fraction, P – Pellet, E1-E5 – elution 1-5

The protein was relatively pure, however there were some contaminants present and based on the $A_{260/280}$ ratio (~1.7), there was also a substantial DNA/RNA contamination (Figure 6.4.1). The eluted protein was dialysed overnight into low salt buffer for cation exchange. In the morning considerable precipitation was observed in the dialysis bag. As the measured $A_{260/280}$ ratio in the supernatant had decreased, the precipitation was believed to be mostly DNA/RNA and the purification was continued, though the presence of the MeCP2 minigene protein in the soluble fraction was not confirmed by an SDS-PAGE analysis.

The filtered solution was applied to a cation-exchange column (ResourceS, GE Life Sciences) and bound protein eluted with a linear gradient of NaCl (50-1000 mM). The resulting fractions were analysed on an SDS-PAGE gel. As seen on Figure 6.4.2 and Figure 6.4.3, the MeCP2 minigene protein bound to the column and eluted around 27 mS/cm. The eluted fractions also contained contaminants and/or degradation products. The flow-through fractions (1-8) contained contaminating proteins and possibly remaining DNA. The $A_{260/280}$ ratio of the pooled fractions was 0.59. After the cation-exchange step the MeCP2 minigene protein was still impure, however many contaminants were separated during this purification step.

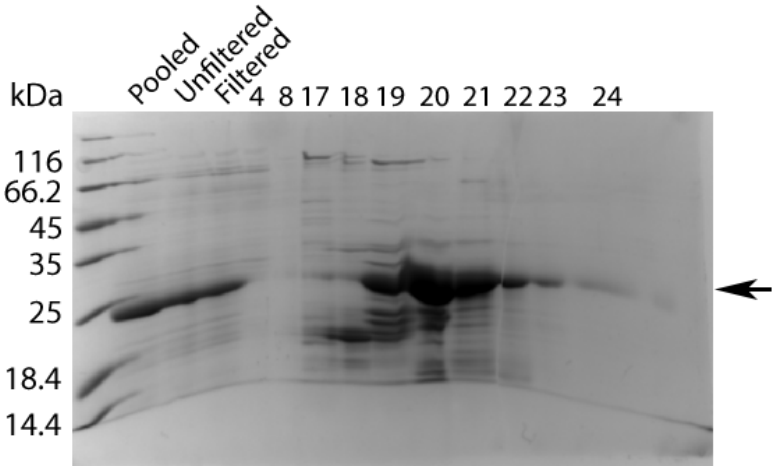

Figure 6.4.2. SDS-PAGE gel of the cation exchange step using 1 ml Resource S column. The filtered overnight dialysed sample was the input. The MeCP2 minigene is shown with a black arrow.

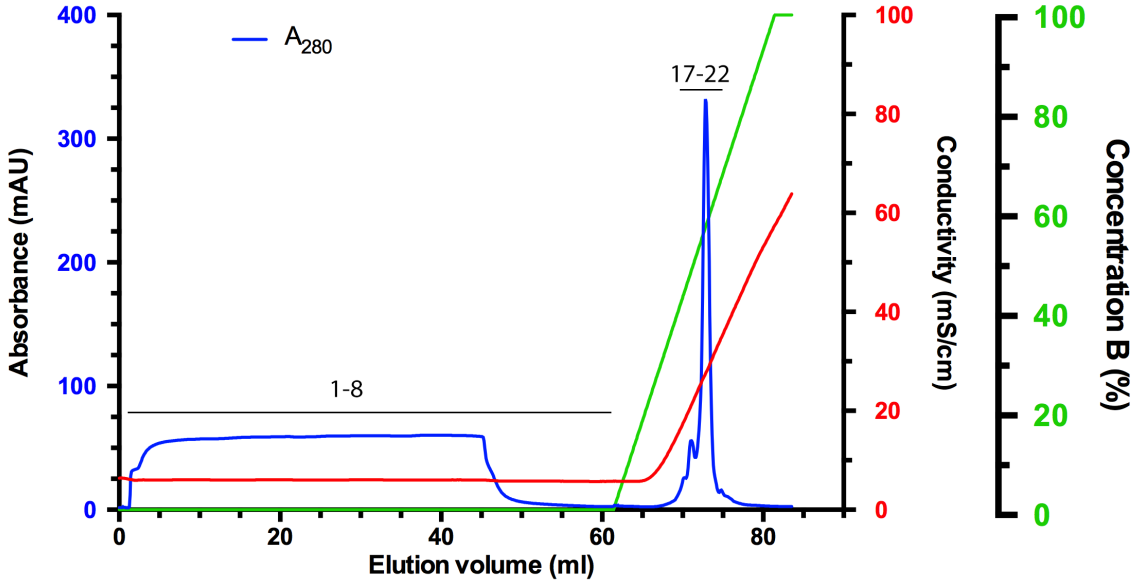

Figure 6.4.3. Cation-exchange chromatogram using 1 ml ResourceS column.

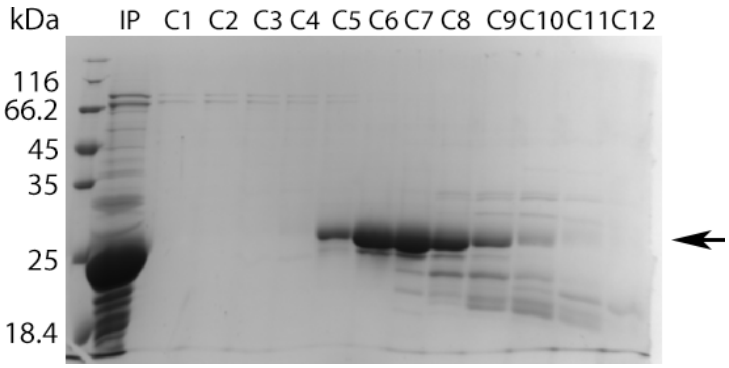

Figure 6.4.4. SDS-PAGE gel showing fractions from size-exclusion chromatography of MeCP2 minigene using Superdex 200 10/300 GL column. MeCP2 minigene is marked with a black arrow.

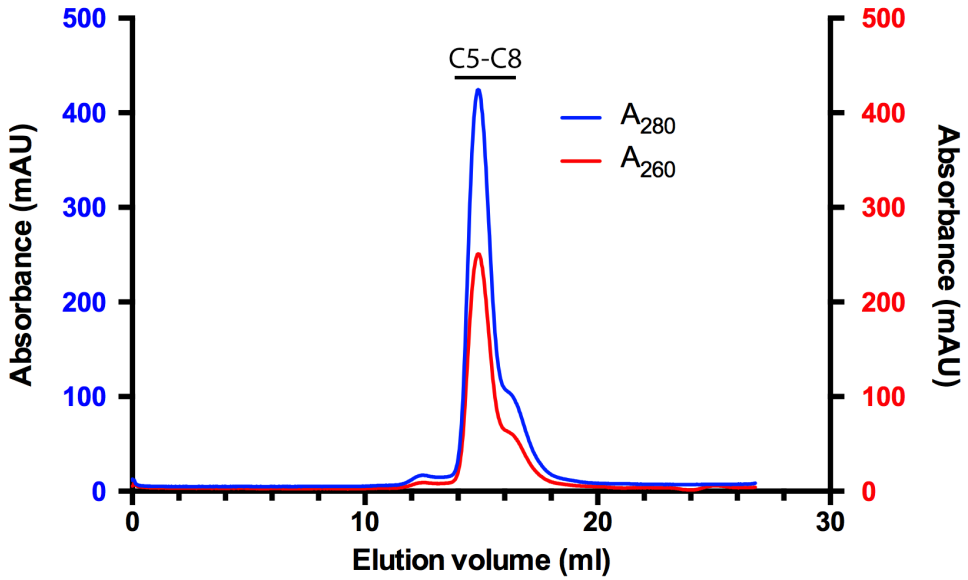

Figure 6.4.5. Size-exclusion chromatogram of MeCP2 minigene using Superdex 200 10/300 GL column.

The resulting fractions that mostly contained the putative MeCP2 minigene were pooled and subjected to size-exclusion chromatography. Some of the protein co-eluted with multiple contaminants, mainly in fractions C8 and C9 (Figure 6.4.4). This was consistent with the chromatogram in Figure 6.4.5 showing an extra peak after the major MeCP2 peak (C5-C8). It also showed that there were no large aggregates as indicated by a lack of a peak near the void volume (~7 ml). Fractions C5-C7 were pooled and the remaining fractions containing the MeCP2 minigene were discarded due to the presence of other proteins and/or degradation products. MALDI-TOF-TOF analysis of the putative MeCP2 minigene bands was used to confirm its identity with 48.7% coverage.

6.5 Purification optimisation

The initial purification protocol produced pure MeCP2 minigene at the expense of final yield. To recover more of the MeCP2 protein at high purity, several changes were made to the purification protocol. First, the salt concentration of the elution buffer in cation-exchange step was reduced from 1000 mM to 500 mM. This allowed the use the same length gradient with better resolution of peaks. In addition, after binding of the MeCP2 to the column, the column was washed with 30% buffer B (same composition as before) to remove contaminants, after which the minigene was eluted with a linear gradient of 30%-100% buffer B. As a result, the MeCP2 minigene was much purer after the cation-exchange step, although some degradation products still co-eluted with the protein (Figure 6.5.1 and Figure 6.5.2).

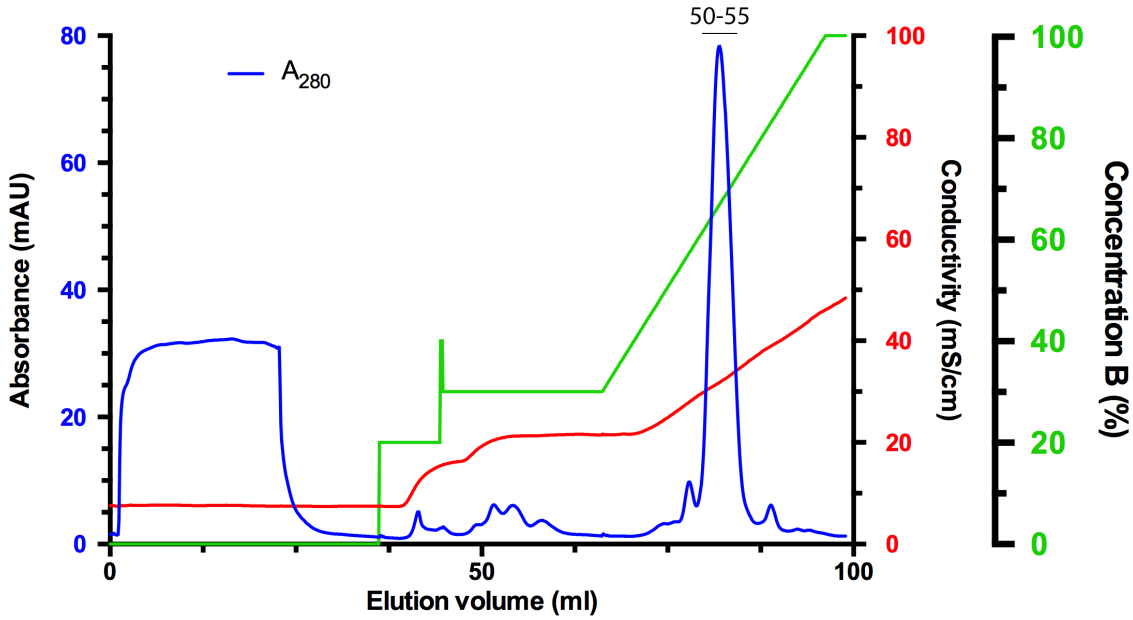

Figure 6.5.1. Cation-exchange chromatogram using 1 ml ResourceS column. The gradient elution was preceded by step elutions of 20% and 30% buffer B.

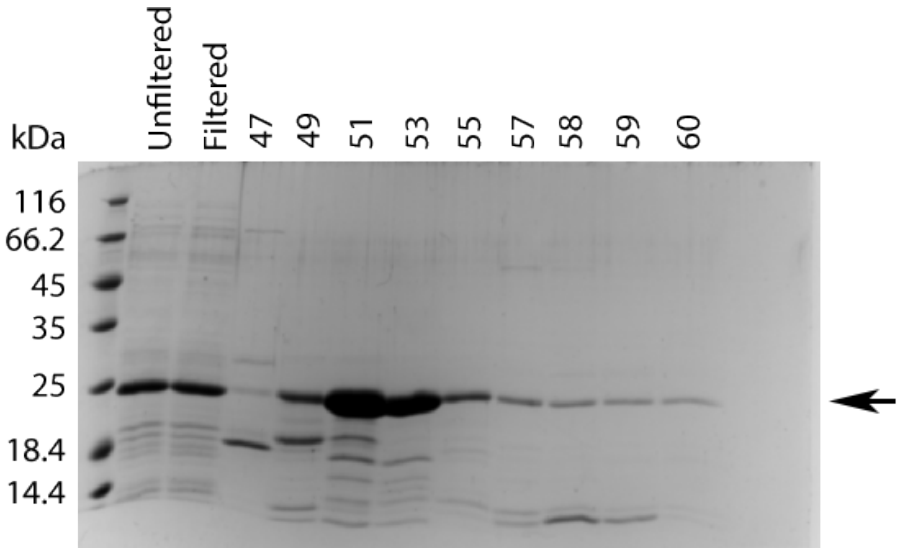

Figure 6.5.2. SDS-PAGE gel of the cation exchange step using a Resource S column. The filtered sample was injected on the column. The MeCP2 minigene is marked with a black arrow.

After the improved cation-exchange chromatography step, size-exclusion chromatography produced protein with much higher purity (Figure 6.5.3 and Figure 6.5.4). Again degradation products were present, but the majority of the sample was the intact minigene protein. In this preparation, the $A_{260/280}$ ratio was 0.79, while most preparations had final values were near 0.56. This improved protocol reliably produced 2-3 mg of MeCP2 minigene protein per 10 g of wet cell pellet to ~95% purity.

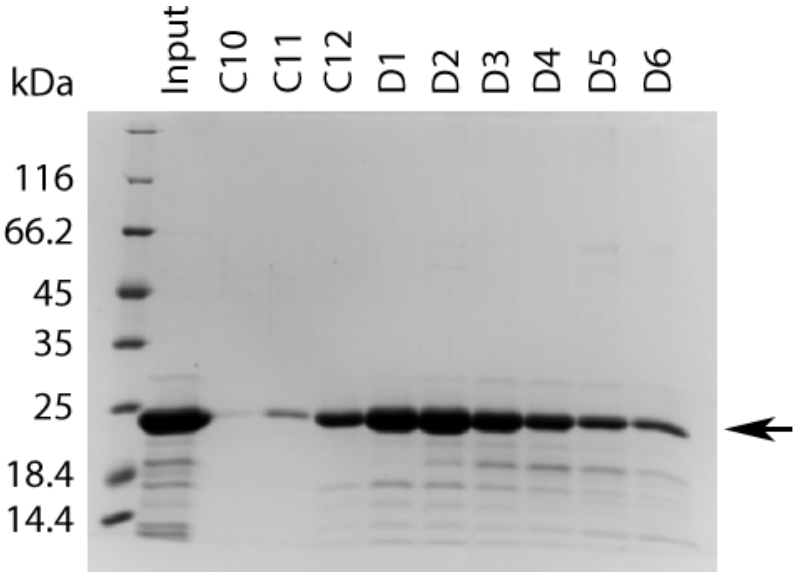

Figure 6.5.3. SDS-PAGE gel of fractions from size-exclusion chromatography of MeCP2 minigene protein. MeCP2 minigene is marked with a black arrow.

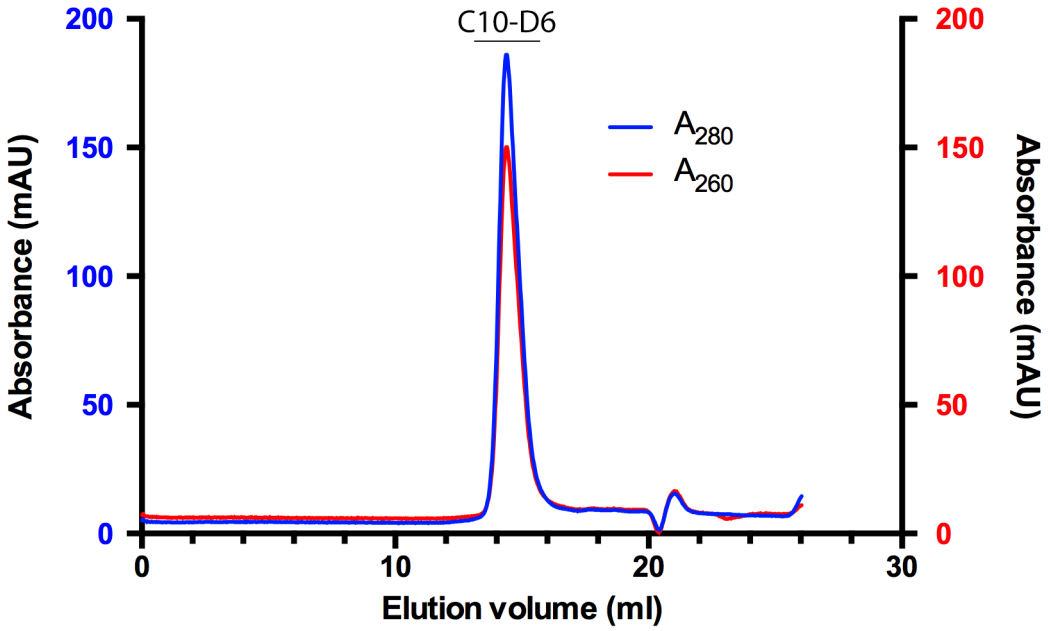

Figure 6.5.4. Size-exclusion chromatogram of MeCP2 minigene using Superdex 200 10/300 GL column.

6.6 Biophysical assays

Unless otherwise stated, the assays in this section were carried out by Ceitidh Taylor and Ciarán Daly under my supervision.

6.6.1 Measuring the stability of the MeCP2 minigene using thermal denaturation assays

To assess whether the MeCP2 minigene was stable and whether it had tertiary structure, a thermal denaturation assay was performed. A series of protein concentrations were screened to see which one would give a good signal. As exemplified in Figure 6.6.1 and Figure 6.6.2, a relatively high concentration of protein was necessary for a good signal from which melting temperature could be obtained.

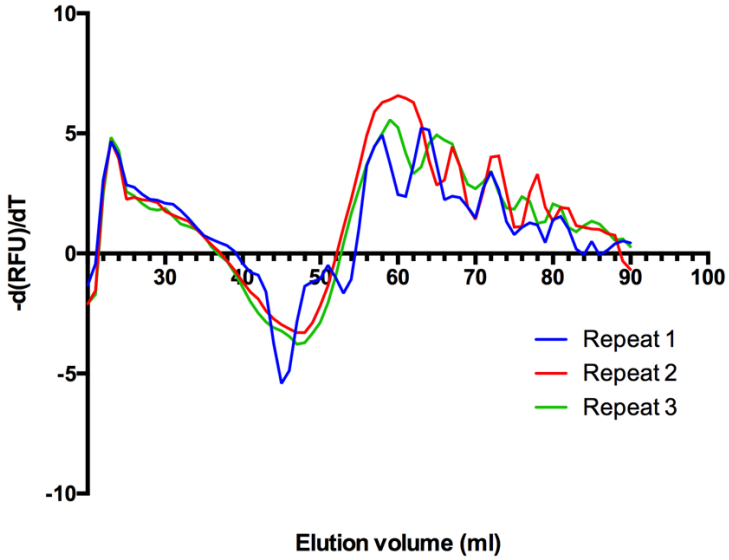

Figure 6.6.1. Thermal denaturation assay of 5 µM MeCP2 minigene protein.

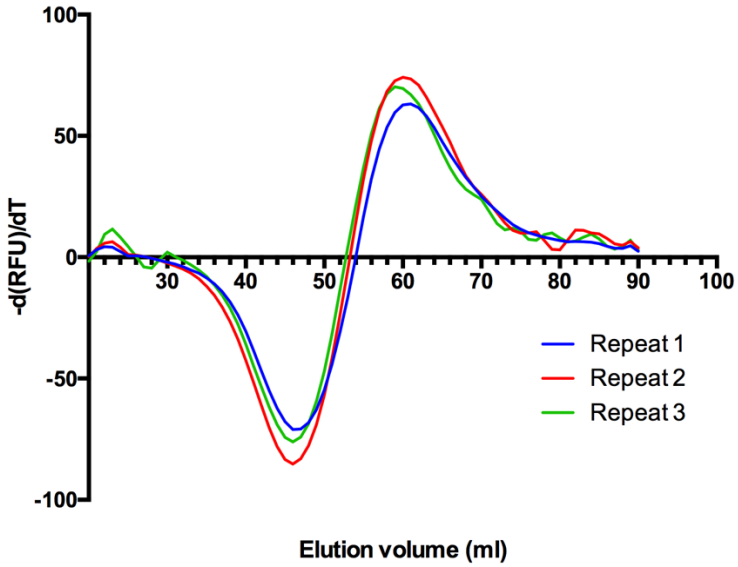

Figure 6.6.2. Thermal denaturation assay of 27 µM MeCP2 minigene protein.

The increase in the fluorescence intensity (a dip on a –d(RFU)/dT graph) shows a protein with tertiary structure undergoing thermal melting. While this assay does not indicate which part of the minigene protein is structured, the melting temperature most likely represents the MBD as the MBD is only 40% unstructured while the whole TRD (spanning residues 198-305) is reported to be more than 80% unstructured, based on circular dichroism analysis (Adams et al., 2007). From the graph the melting temperature of the minigene was found to be between 46˚C and 47˚C (Figure 6.6.2).

6.6.2 MeCP2 minigene has a functional MBD

While the thermal denaturation assay showed the presence of some tertiary structure, it did not confirm whether the MBD was functional. In order to assess this, the protein was assayed using electrophoretic mobility shift assays (EMSA). This assay tracks the migration of a species through a porous medium, with the migration speed being dependent on the size of the species. Larger species move more slowly through the medium and are therefore shifted to a lesser extent than smaller species. As interacting partners are added, the size of the complex increases and the speed of migration decreases.

EMSA was used to detect whether MeCP2 minigene protein bound to a fluorescent DNA probe that contained a symmetrically methylated CpG dinucleotide on both strands (Table 6.6.1). This probe was the same as the one used in the crystallisation of the MeCP2 MBD (Ho et al., 2008). Also included were the wild-type MeCP2 MBD, containing residues 1-205, as a positive control; and as a negative control a R111G 1-205 construct. The R111G mutant is able to bind DNA but does not have specificity for methylated DNA. A negative control for MBD that would not bind DNA was not available. Initially no competitor was used, as seen in Figure 6.6.3. There was a weak shift when using 1 µM of MeCP2 minigene protein. However, both the positive control (wild-type MeCP2 MBD) and the negative (R111G MBD) (both proteins received from Rebekah Tillotson) also bound the DNA, apparently to an equal extent. This indicated that there were interactions with the DNA backbone irrespective of the presence of the methyl group.

Table 6.6.1. Fluorescent DNA probe used for the EMSAs. The forward sequence contained the DY-681 fluorophore at its 5' terminus which has an emission maximum of 708 nm. The character "5" indicates a 5-methyl C base.

Forward sequence	5'DY-681 – ATA GAA GAA TTC 5GT TCC AG 3'
Reverse sequence	3' AT CTT CTT AAG G5A AGG TCT 5'

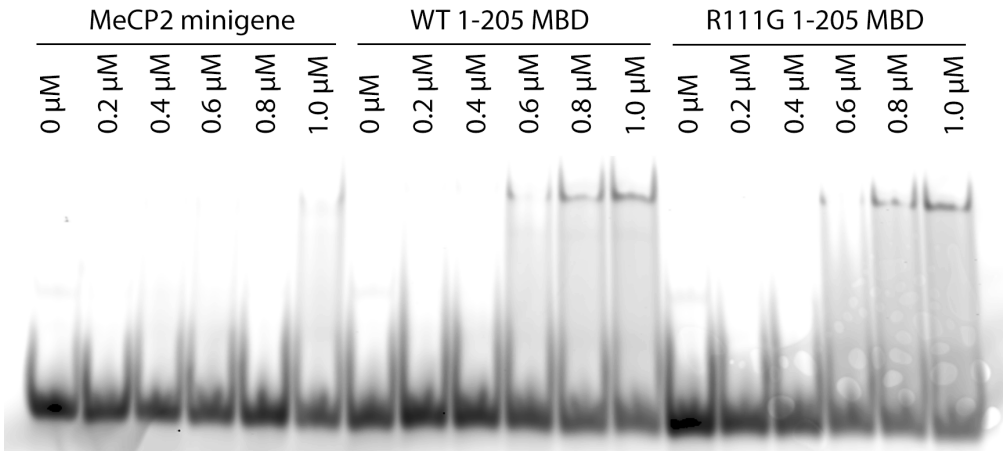

Figure 6.6.3. EMSA using 8% acrylamide native gel to detect whether MeCP2 minigene has a functional MBD. The amount of labeled DNA was 0.5 µM in all samples.

To be able to detect binding specific to methylated DNA, an excess of unlabelled dA/dT competitor was used. The amount of DNA was also lowered to 0.1 µM in the binding reaction as using 0.5 µM of the labelled DNA caused saturation of the scanner. However, as seen in Figure 6.6.4, using 1.5 µg of unlabelled competitor completely abolished the binding of the fluorescent probe to any of the tested proteins.

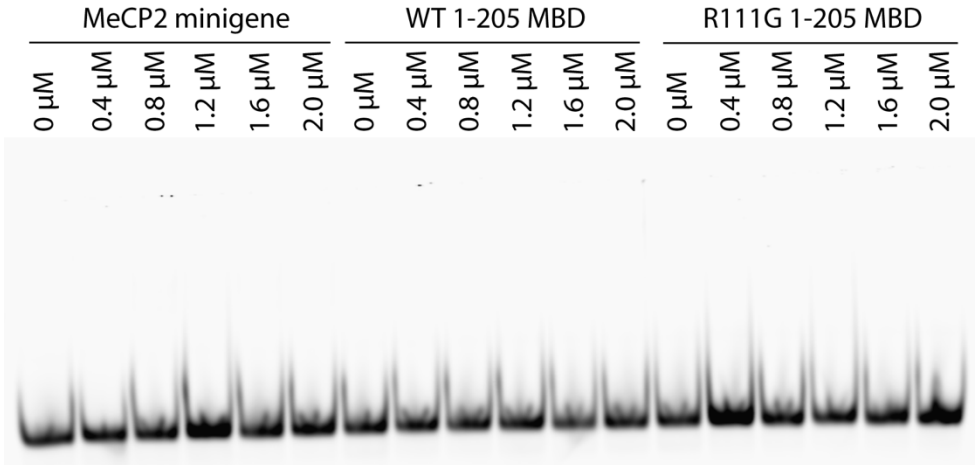

Figure 6.6.4. EMSA using 8% acrylamide native gel and 1.5 µg of unlabelled dA/dT competitor. The amount of labelled DNA used was 0.1 µM.

The amount of competitor was lowered to 500 ng and the assay was repeated with a wider range of MeCP2 minigene concentrations (Figure 6.6.5). With lower competitor concentration, the fluorescent DNA was bound to all three proteins, and there were probable multiple binding events of MeCP2 to DNA as evidenced by the higher molecular weight band in the 3 µM WT MBD lane. The protein in this lane was not stuck in the well, but had not migrated far from the start of the well. However again it was impossible to definitely say

whether the minigene protein binds methylated DNA specifically or if the DNA binding is due to nonspecific contacts with the DNA phosphate backbone. Analysis of the literature showed that this was likely due to the length of the DNA fragment used, as 23 base pair long fragment of DNA has been shown to have only marginally higher affinity for methylated DNA over unmethylated DNA (Ghosh et al., 2008).

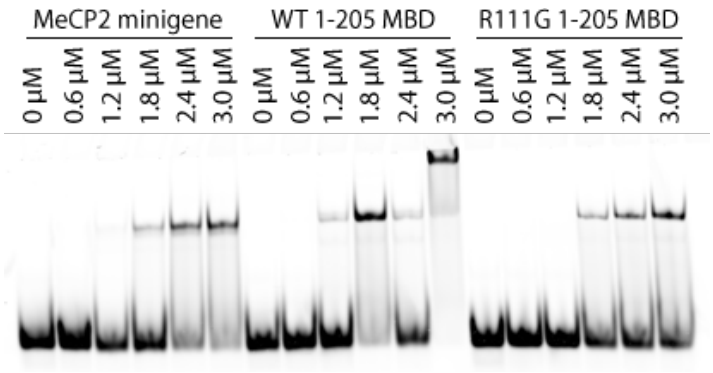

Figure 6.6.5. EMSA using 8% acrylamide native gel and 500 ng of unlabelled dA/dT competitor per well. The amount of labelled DNA used was 0.1 µM.

6.6.3 Does MeCP2 display increased affinity towards DNA in the presence of TBL1XR1?

The initial EMSAs showed that there was binding of MeCP2 minigene to DNA at concentrations between 1.2 and 3.0 µM, although specific binding to methylated DNA could not be established. Nevertheless, the effect of TBL1XR1 on this interaction was examined using supershift assays. This is an EMSA with a third interacting component that should produce an additional band on the gel, which is called a supershift. Initially the assays were carried out with increasing concentrations of TBL1XR1 in the presence of 5 µM MeCP2 and no competitor to see whether binding to DNA still occurred (Figure 6.6.6).

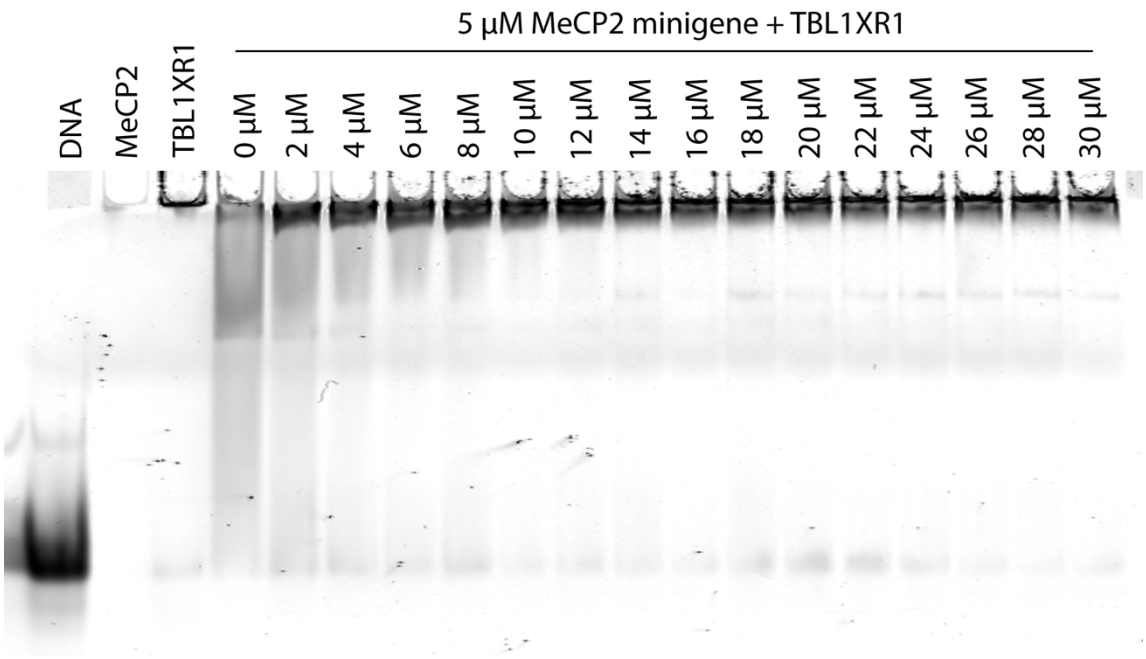

Figure 6.6.6. EMSA using 8% acrylamide native gel and no dA/dT competitor. Increasing concentrations of TBL1XR1 in the presence of 5 µM MeCP2 minigene were probed. The amount of labelled DNA used was 0.1 µM. MeCP2 lane (second) contained 5 µM protein. TBL1XR1 lane (third) contained 20 µM protein and 0.1 µM DNA.

Unfortunately, a complex was formed that migrated only marginally into the gel, with the majority of the MeCP2-DNA-TBL1XR1 being stuck in the wells. This could be due to the complex being too large to enter the gel. Alternatively, the overall charge of the complex could have been reduced upon complex formation, which would have affected its migration in the electric field. There was also an issue with background bands, and the presence of an unknown band in the DNA sample. The polyacrylamide gel was changed from 8% to 4%, and competitor was used again in hopes that it would help produce discrete bands instead of the smeared bands present in the protein titration lanes. Furthermore, during the course of these experiments the K_D values for MeCP2 NID and TBL1XR1 interaction were established from the SPR and FA assays (Chapter 5) to be in the range of 10-20 µM, therefore an excess of TBL1XR1 (40 µM) was added to the reaction mix to have a majority of MeCP2 minigene protein bound to TBL1XR1, and MeCP2 minigene was titrated instead. In this case, the supershifted band was observed in a lane with 3 µM MeCP2 minigene and 40 µM TBL1XR1, and possibly with 2 µM MeCP2 (Figure 6.6.7). The complex had also migrated into the gel this time and did not remain in the wells. The DNA was shifted at lower concentrations of MeCP2 in the presence of TBLXR1, compared to MeCP2 on its own. This hinted at the possibility that there is enhanced binding of MeCP2 to DNA upon TBL1XR1 binding, although there remained

a possibility that the effect seen was an artefact caused by the high concentration of TBL1XR1 used in the reaction, which could have increased the local concentration of the MeCP2 and DNA due to molecular crowding.

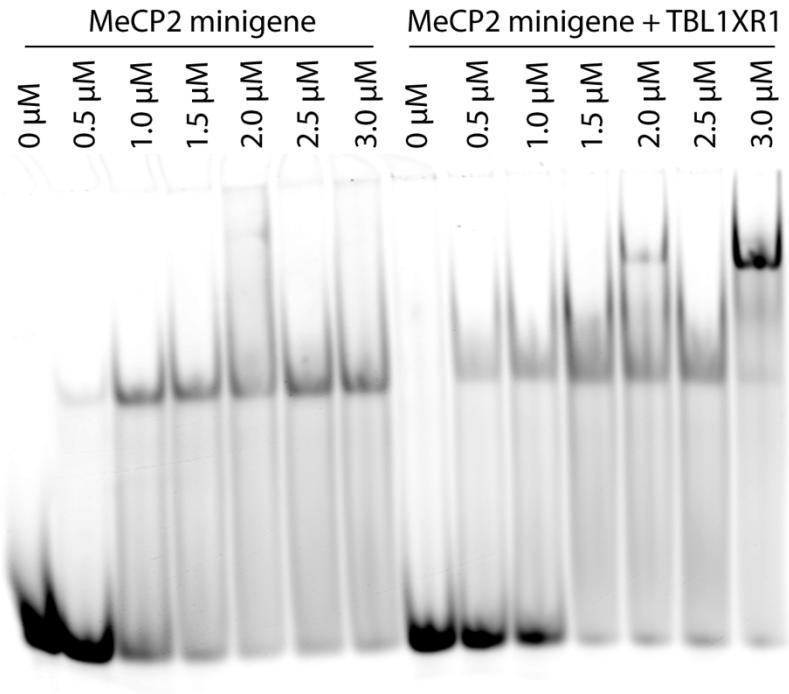

Figure 6.6.7. EMSA using 4% acrylamide native gel and 500 ng of dA/dT competitor. The concentrations refer to MeCP2 minigene. The amount of labelled DNA used was 0.1 µM. The amount of TBL1XR1 in supershift lanes was 40 µM.

While the supershift assays were not conclusive, the results warranted further investigation. EMSAs are good for observing binding, including multiple binding events as seen in supershift assays, as well as estimating K_D values. However, their limitation is that during the assay the components are not at chemical equilibrium as products get separated from reactants. Additionally, in this case quantifying DNA binding was difficult due to presence of smeared bands. To obtain an equilibrium measurement, FA assays were used. To determine whether MeCP2 exhibited increased affinity for DNA when TBL1X or TBL1XR1 was bound, the affinity of the MeCP2 minigene protein for the methylated probe used had to be calculated first. The probe used was the same as in Table 6.6.1, except the fluorophore was Cy3 to match the fluorimeter specifications. Using serial dilutions of MeCP2 minigene protein ranging from 5 µM to 12.5 nM produced a calculated K_D value of 64 ± 21 nM, which was in the expected range for the MBD based on previous literature (Figure 6.6.8) (Fraga et al., 2003). However the measured anisotropy exceeded 0.4, which, assuming random distribution of molecules

in the solution, indicated scattered light (Lakowicz, 2006). This polarised light in the assay would interfere with the readings and would skew the calculated K_D values.

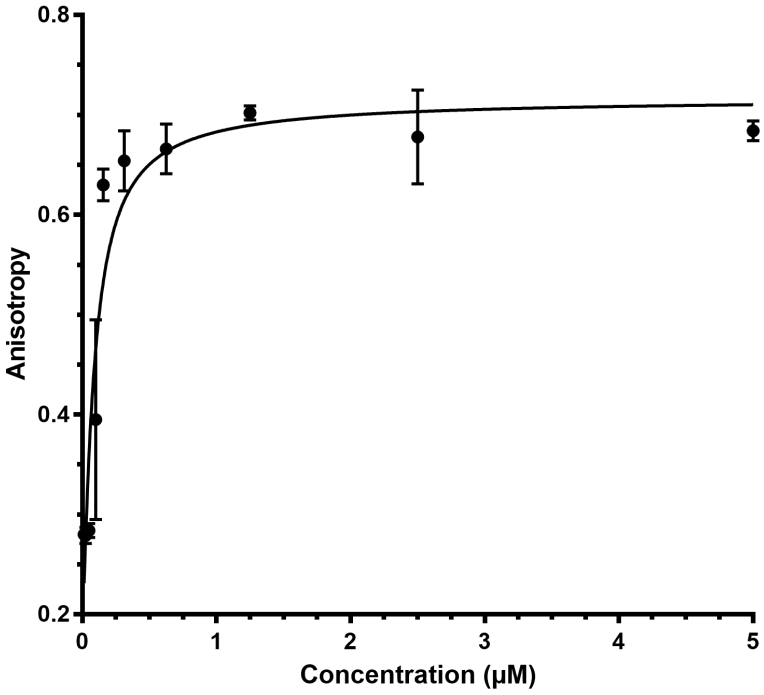

Figure 6.6.8. Fluorescence anisotropy assay using Cy3-tagged DNA probe (45.4 nM) and increasing concentration of MeCP2 minigene protein. Excitation and emission wavelengths were 530 nm and 562 nm, respectively. All experiments were done in triplicate. The binding curve was fitted using non-linear regression. Error bars are one standard deviation.

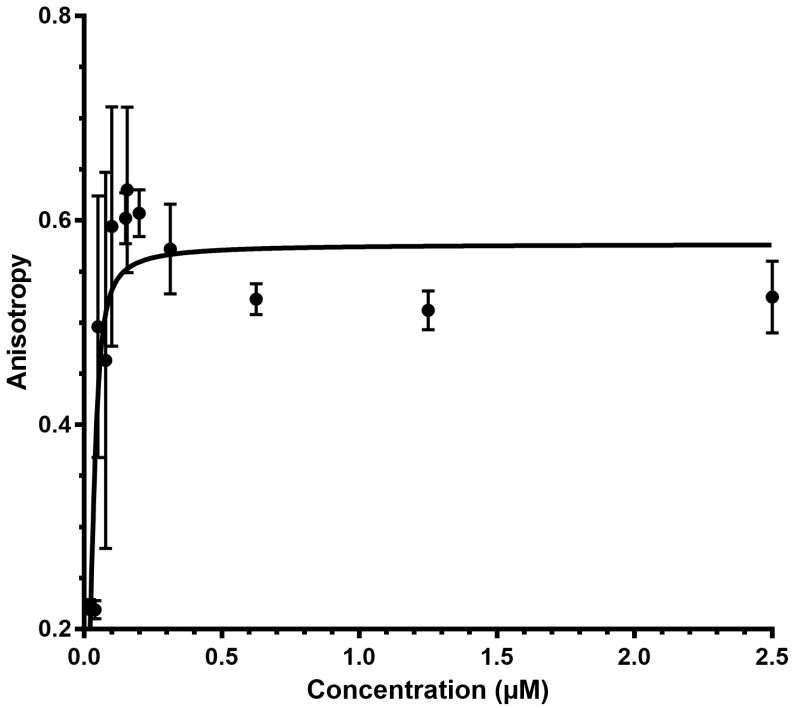

Figure 6.6.9. Fluorescence anisotropy assay using Cy3-tagged DNA probe (100 nM) and increasing concentration of MeCP2 minigene. Excitation and emission wavelengths were 530 nm and 562 nm, respectively. All experiments were done in triplicate. Binding curve was calculated using non-linear regression. Error bars are one standard deviation.

Another assay was set up to see how consistent the assay was between runs, and to understand the origin of the scattered light. The concentration of the probe was also increased from 45.4 nM to 100 nM to boost the fluorescence signal (Figure 6.6.9). This time the measurements between wells varied immensely in the middle of the concentration series. It was noted that at those protein concentrations the meniscus in the wells was very pronounced. This likely affected the measured values due to changes in the reflection and refraction light intensities. In addition, the measured anisotropy values were again above 0.4, still indicating presence of scattered light.

No K_D values could be calculated from that experiment as the large standard deviations made curve fitting highly unreliable. To help with the scattering of the light by aggregates in the solution, as well as with effect of the meniscus, this assay was repeated with 0.17% w/v of Tween 20, a nonionic detergent. It was observed that the pronounced meniscus seen previously disappeared with the addition of Tween 20 (Figure 6.6.10). It also helped with the aggregation, as seen from the anisotropy values which now spanned sensible values i.e. close to the expected values from our calculations. Unfortunately, the addition of Tween 20 had a

pronounced effect on the measured parallel light intensities. This was thought to be due to the formation of micelles, which could scatter light. While this background signal could be subtracted from the experimental values, the micelles might interfere with the assay by sequestering either the probe or the protein and producing erroneous anisotropy values.

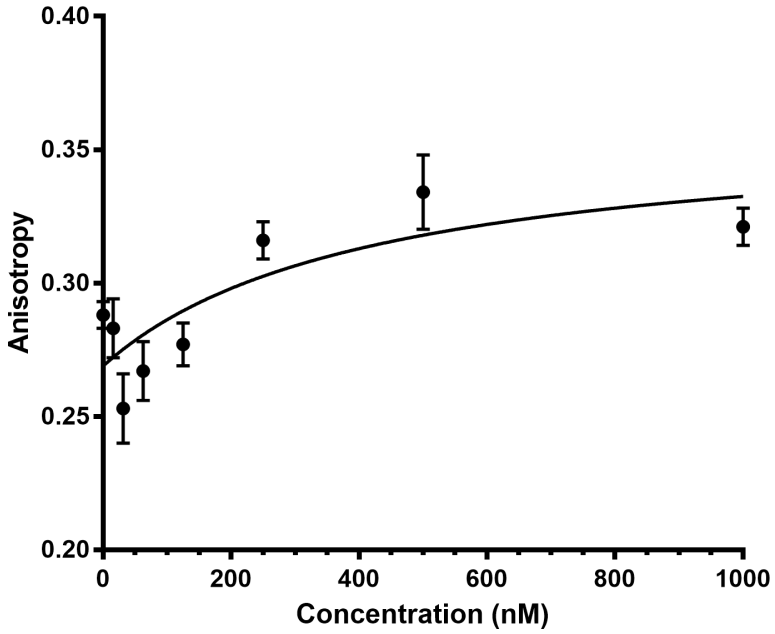

Figure 6.6.10. Fluorescence anisotropy assay using Cy3-tagged DNA probe (100 nM) with 0.17% Tween 20, and increasing concentration of MeCP2 minigene protein. Excitation and emission wavelengths were 530 nm and 562 nm, respectively. All experiments were done in triplicate. Binding curve was fitted using non-linear regression. Error bars are one standard deviation.

The assay was repeated using Tween 20 below its critical micelle concentration (CMC) to prevent the formation of micelles (Figure 6.6.11). The graph showed that there was an increase in anisotropy upon the addition of MeCP2 minigene up to 250 nM, after which the anisotropy started dropping. When the rata data was examined, the total intensity of the measured light dropped at those intensities as well. This indicated that there was quenching of the fluorophore upon binding to MeCP2 minigene. In fact, in Figure 6.6.8 and Figure 6.6.9 the same effect was seen to a smaller extent, probably masked by the scattered light. In Figure 6.6.8 the effect was very small due to already high measured anisotropy values. However, due to the high standard deviation values in Figure 6.6.9 the quenching was not spotted because it was thought the intermediate protein concentration anisotropy values were completely unreliable and likely resided below the maximum values measured. In

addition, the problems with the meniscus meant the total intensity values produced were rather low and variations in the data were not apparent.

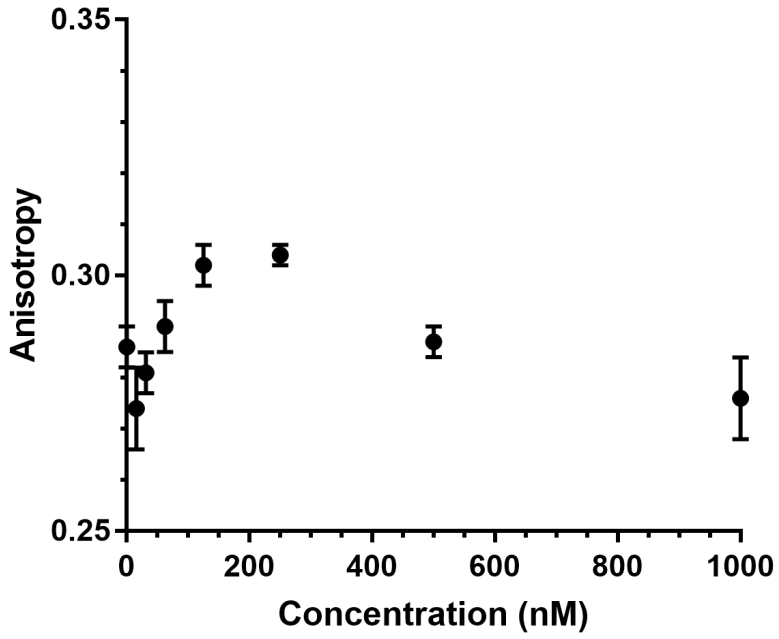

Figure 6.6.11. Fluorescence anisotropy assay using Cy3-tagged DNA probe (100 nM) with 0.005% Tween 20, and increasing concentration of MeCP2 minigene. Excitation and emission wavelengths were 530 nm and 562 nm, respectively. All experiments were done in triplicate. Error bars are one standard deviation.

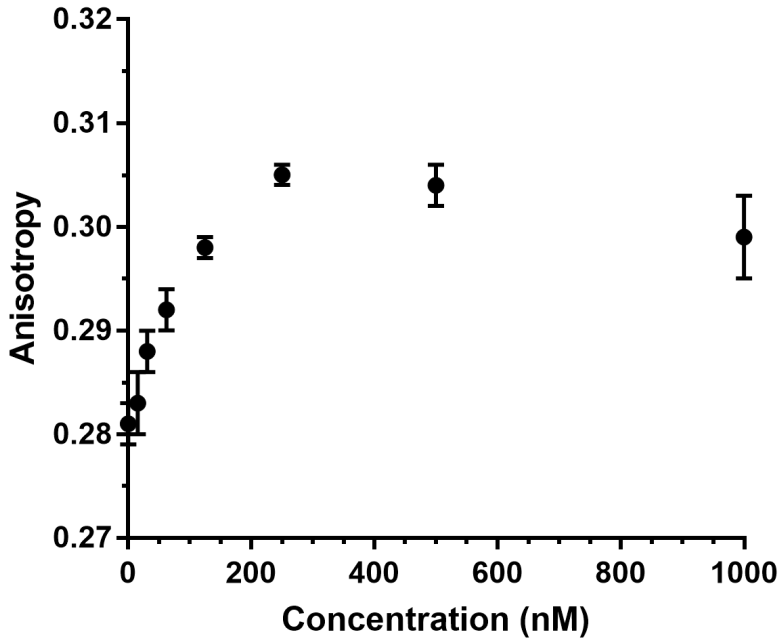

Figure 6.6.12. Fluorescence anisotropy assay using Cy3-tagged DNA probe (100 nM) with 0.005% Tween 20, and increasing concentration of MeCP2 minigene. Excitation and emission wavelengths were 530 nm and 567 nm, respectively. All experiments were done in triplicate. Binding curve was fitted using non-linear regression. Error bars are one standard deviation.

Because the emission maximum of Cy3 is around 570 nm, the same plate was measured at 567 nm to see whether increasing Stoke's shift would have an effect on the measurements. As shown in Figure 6.6.12, the datapoints followed a similar pattern when measured at 562 nm, with a loss of anisotropy and intensity (not shown) at higher MeCP2 concentrations. Unfortunately, due to time constraints these assays were not optimised further during this project.

6.7 Discussion

Because knock-in mice containing MeCP2 minigene protein showed a mild RTT-like phenotype, there were two possibilities – either the MBD or NID function was somehow perturbed due to the deletion of the conserved N- and C-termini, as well as the internal linker between the MBD and NID, or those deleted regions had some role in RTT pathology. EMSA and FA assays were set up to determine whether the MBD and NID were functional in this construct. In addition, whether MeCP2 displays higher affinity for DNA when it is bound to its interaction partner TBL1XR1 was examined.

The MeCP2 minigene protein was expressed and purified from *E. coli*. Due to the promiscuous nature of the MBD DNA binding, the expression protocol was adapted from Klose *et al.* and the induction time was kept short to minimise the toxic effects of the recombinant MeCP2 protein. A three step protocol consisting of IMAC, followed by cation-exchange chromatography and SEC produced 2-3 milligrams of MeCP2 minigene at over 95% purity per 10 grams of wet cell pellet. First a TDA confirmed that the MeCP2 minigene had some tertiary structure. While this assay probes the whole protein as a whole, rather than being specific to any part of the protein, it was concluded that the tertiary structure most likely belongs to the MBD as the NID is very likely to be unstructured. Analysis of the TDA data showed that MeCP2 minigene protein denatures between 46 and 47˚C.

To assess whether the MBD is functional, several EMSA assays were set up. These assays were unable to distinguish between specific and nonspecific binding of MeCP2 to DNA, although they did show that MeCP2 is able to shift DNA, meaning the MBD is functional and able to make at least non-specific contacts with DNA. The effect of TBL1XR1 on the affinity of MeCP2 minigene to DNA was looked at next. A supershift, i.e. ternary complex consisting of MeCP2-DNA-TBL1XR1, was observed by EMSA. In the same assay, an increase in the affinity of MeCP2 to DNA was indicated as lower concentrations of MeCP2 in the presence of TBL1XR1 were able to bind DNA compared to MeCP2 on its own. The results were not conclusive but hinted at the possibility of MeCP2 displaying a higher affinity for DNA when bound to TBL1XR1, indicating cooperative binding. To further investigate this, a fluorescence anisotropy assay was set up using a fluorescently-labelled DNA and MeCP2 minigene protein. Initial experiments produced a K_D value of 64 ± 21 nM for the interaction between the DNA probe and MeCP2 minigene, although the low signal intensity and presence of scattered light likely makes this value erroneous. As the wells were observed to have a pronounced meniscus upon the addition of MeCP2 protein, a non-ionic detergent (Tween 20) was added at a concentration below its CMC. This lowered the measured anisotropy values to physically sensible values by removing the scattered light. This unmasked another issue, which was quenching of the fluorescence signal at high MeCP2 concentrations. As there was no time to complete these assays, a proper K_D measurement for the MeCP2-DNA interaction could not be established, nor could the effect of MeCP2-TBL1XR1 interaction on the affinity of MeCP2 to DNA be deduced.

7 Conclusions

7.1 TBL1X and TBL1XR1 were successfully expressed, purified, and characterised

During this project both TBL1X and TBL1XR1 were successfully expressed in Sf9 insect cells. Initially the expression was carried out in *E. coli*, however after a lengthy optimisation period it was apparent that it is unfeasible to express these proteins in bacterial system as the yields obtained were in low micrograms per 20 grams of wet cell pellet mass. After switching to insect cell expression system, the expression of various TBL1X and TBL1XR1 constructs regularly reached 4-5 mg of protein per 20 g of wet cell pellet.

Successful purification protocols were established for TBL1X and TBL1XR1 WD40 domain constructs. At first there were issues with these two proteins during ion-exchange steps. After multiple unsuccessful trials to overcome problems like aggregation and non-binding that occurred during either cation or anion exchange step, this purification step was cut out of the protocol. However, it was possible to purify these two proteins to high purity using IMAC and SEC. The established protocol regularly produced protein that was over 95% pure with no contaminating DNA or RNA as observed by A_{260}/A_{280} ratio.

Both TBL1X and TBL1XR1 were characterised biochemically and biophysically. It was shown that the purified recombinant protein behaves as expected *in vitro* by interacting with wild-type MeCP2 NID peptides, but fails to bind peptides with known RTT mutants K304E, K305R, and R306C, as well as interacting more weakly with MeCP2 NID peptide truncated at amino acid 308. Knock-in mice bearing an MeCP2 mutation that truncates the protein at amino acid 308 have been shown to have RTT syndrome, consistent with the observations of impaired binding. TBL1X and TBL1XR1 were also characterised using thermal denaturation assays and were found to be relatively thermostable, with melting temperatures around 60˚C. While testing the effect of different cations and anions, it was found that while NaCl did not affect the thermostability of the TBL1X/TBL1XR1 WD40 domain to a large extent, ammonium sulphate had a large stabilising effect. In a similar fashion it was seen that these two proteins

are relatively unaffected by pH values between 6 and 10, while they are greatly destabilised at pH below 6.

7.2 The structure of MeCP2 bound to TBL1XR1 WD40 domain was solved

After an initial sparse-matrix screening and subsequent optimisation of crystallisation conditions, co-crystals of MeCP2 and TBL1XR1 were obtained. While the grown crystals were stacked plates, it was possible to separate them and freeze individual crystals, or in some cases stacked crystals with non-overlapping parts. The best diffraction set collected from these crystals was at 2.5 Å. The structure was solved using molecular replacement using a human TBL1XR1 WD40 domain as a model. The structure showed that there was MeCP2 peptide present bound to two of the four TBL1XR1 molecules in the asymmetric unit. While the MeCP2 peptide spanned residues 285-309, only residues 298-307 were visible in the electron density map, indicating that the rest of the peptide was flexible and most likely not involved in binding.

7.3 MeCP2 and TBL1XR1 interaction was characterised using structural and biophysical methods

7.3.1 MeCP2 interacts with TBLR1 using four key residues

Structural analysis of the MeCP2/TBL1XR1 co-crystal revealed several important facts. The most important finding was that all four known RTT mutations that occur in the NID (P302R, K304E, K305R, and R306C) make extensive contacts with TBL1XR1 WD40 domain. Equally importantly, it could be explained, on a structural level, why these mutations disrupt the binding of the MeCP2 to TBL1XR1, and in turn, cause RTT. These mutations prevent binding by disruption of the extensive hydrogen bonding network between MeCP2 and TBL1XR1, disruption of local electrostatics, and through steric clashes.

As predicted from the biochemical nature of the MeCP2 NID, as well as from the surface of TBL1XR1 where MeCP2 was thought to bind, MeCP2-TBL1XR1 interaction was shown to be salt-sensitive. MeCP2 NID peptide binding to TBL1XR1 was abolished around 400 mM NaCl, and around 200 mM AmSO4. This information had been used previously in crystallisation trials by excluding screens with conditions that had high levels of salts in them.

The importance of various TBL1XR1 residues around the MeCP2 binding site were assessed through mutagenesis. While it was not feasible to test all of the amino acids, a number of amino acids that were seen to be involved in binding were mutated to either similar amino acids (aspartate to asparagine, glutamate to glutamine, tyrosine to phenylalanine) or to neutral (alanine) ones. The effects of these mutations were assessed using surface plasmon resonance assays to measure binding constants. True to the predictions made based on the structure, all but two of the amino acids, when mutated, either abolished or reduced the binding MeCP2 to TBL1XR1. The reduction in the binding constants was large enough to abolish the interaction in a cellular context, as seen through the co-localisation assays (Matthew Lyst, personal communication). Two of the mutations, $TBL1XR1^{E351D}$ and $TBL1XR1^{Y446F}$, showed similar levels of MeCP2 binding to wild-type TBL1XR1. E351D mutation might change the local geometry of the binding site but maintains the negative charge, which is likely to be important for peptide binding, while Y446F is unable to form a hydrogen bond with the MeCP2 peptide, as seen in slight reduction in K_D values. The obtained K_D values for the wild-type protein were confirmed by using fluorescence anisotropy assays, which resulted in similar dissociation constant values as those obtained from SPR. In the process of optimising the biophysical assays to measure the various parameters of these interactions, these assays were also optimised for the use of medium- and high-throughput screening of small drug-like molecules, which will be an important future direction for this project.

The exact overlap between the four NID mutations that cause RTT, and the four amino acids that are most important for MeCP2 binding demonstrated with high confidence that the main function of MeCP2 is to bind TBLR1, that NID mutations cause the loss of binding to NCoR/SMRT, and that this is sufficient to cause RTT. Furthermore, this also casts doubt on the competing hypothesis that the primary function of the NID is to bind DNA, as mice harbouring TBLR1 mutations that abolish MeCP2 binding nevertheless display neurological symptoms, despite retaining wild-type MeCP2 (Matthew Lyst, Jim Selfridge, personal communication). Furthermore, no RTT mutations that retain the electropositive nature of the NID (such as K305R) have been demonstrated to disrupt the NID binding to DNA, hinting at the possibility that the DNA binding is non-specific owing to the highly electropositive nature of the NID. Taken together, the data presented in this thesis strongly support the hypothesis that MeCP2 acts as a molecular linker between DNA and NCoR/SMRT co-repressor complex.

7.3.2 MeCP2 binding shows similarities with other WD40 domain peptide interactors

As the individual WD40 proteins are able to bind to different types of ligands, there is no defined binding pocket or motif in these domains between different structures owing to the variable biochemical nature of their binders. Only the binding site locations are shared between different complexes (top, bottom, or circumference of the WD40 domain). However, there are some general features that occur in these different complexes.

In most of the structures, the peptides are bound near the entrance of the central channel of the WD40 domain. These peptides usually use the loops between the beta strands to bind to the top face of the WD40 domain, although this is not always the case, as for example colicin E9 bound to TolB (PDB ID: 2IVZ) makes extensive use of the amino acids that form the core beta sheets of the WD40 domain (Figure 7.3.1). In addition, some of the peptides can insert (usually) one amino acid into the central channel. This is used, for example, for MLL1 recognition by WDR5 (Figure 7.3.2). Furthermore, the mode of recognition of these peptides is not conserved and can vary from π-π and cation-π interactions to hydrogen bonding and salt bridges.

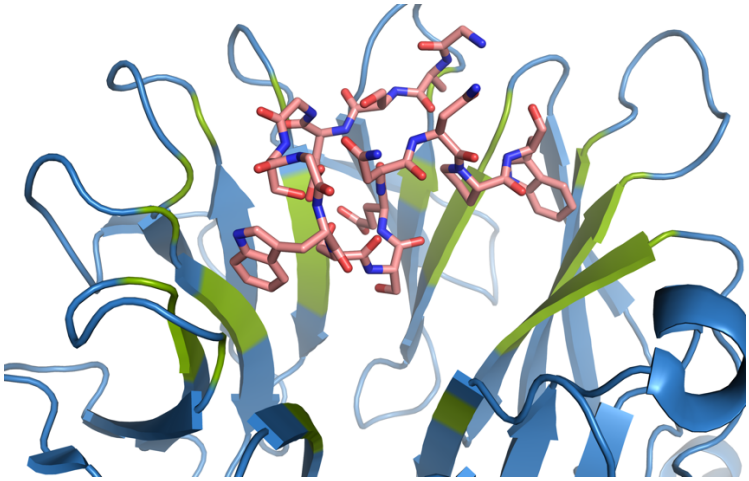

Figure 7.3.1. Binding of colicin E9 (light pink) to TolB (dark blue) (PDB ID: 2IVZ). Residues that interact with colicin E9 peptide have been coloured green. For clarity, these are only shown in cartoon representation.

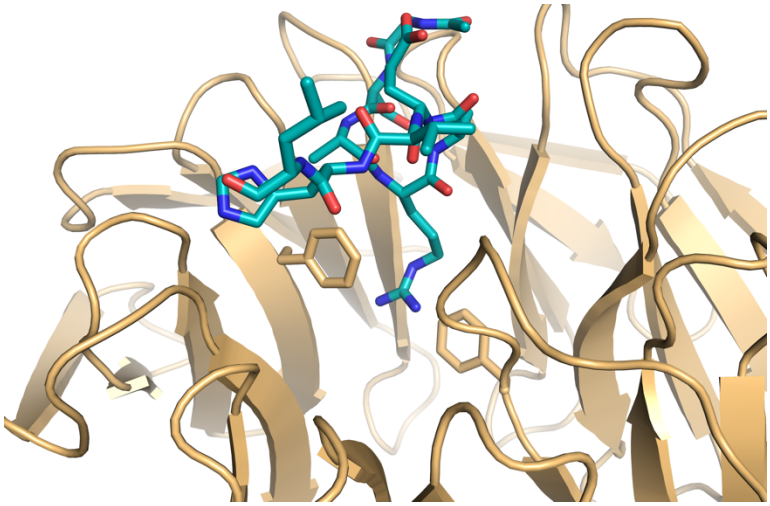

Figure 7.3.2. Binding of MLL1 Win motif (teal) to WDR5 WD40 domain (light orange) (PDB ID: 3EG6). The arginine is inserted into the central channel of the WD40 domain where it forms cation-π interactions with two phenylalanines.

In the case of MeCP2, it is the lysine 304 (K304) that is inserted into the central channel to form hydrogen bonds and salt bridges with TBLR1. Interestingly, rather than utilising side chains of the WD40 domain for recognition, like for example MLL1 shown above, K304 binds to backbone carbonyl groups of TBLR1 residues, leaving the side chains free to engage in WD40 blade-blade packing. In addition, all residues that MeCP2 utilises to directly contact TBLR1 are found in the loops. Finally, MeCP2 utilises a mixture of different bond types for binding, just like many of these peptides.

The main chain and side chain intramolecular hydrogen bonds in MeCP2 were also observed in many other peptides (Figure 7.3.3). It is unknown whether these bonds exist in the unbound state, however these interacting peptides are very often predicted to be in intrinsically unstructured regions. It is therefore likely that they help the peptides adopt a proper shape required for binding to these channels, suggesting that the stabilisation of these peptide folds is critical to their binding to the central channel. In the absence of these bonds, the loss of entropy upon binding of MeCP2 to TBLR1 would not be compensated by the bonds made to the WD40 domain. However, this is not always the case. For example, histone H3, which inserts an arginine into the central channel, also binds across the top face of the WD40 domain of Nurf55 and is unable to form most of these intramolecular contacts (Figure 7.3.4). In this case, the increased binding surface area likely compensates for the loss of these stabilising bonds.

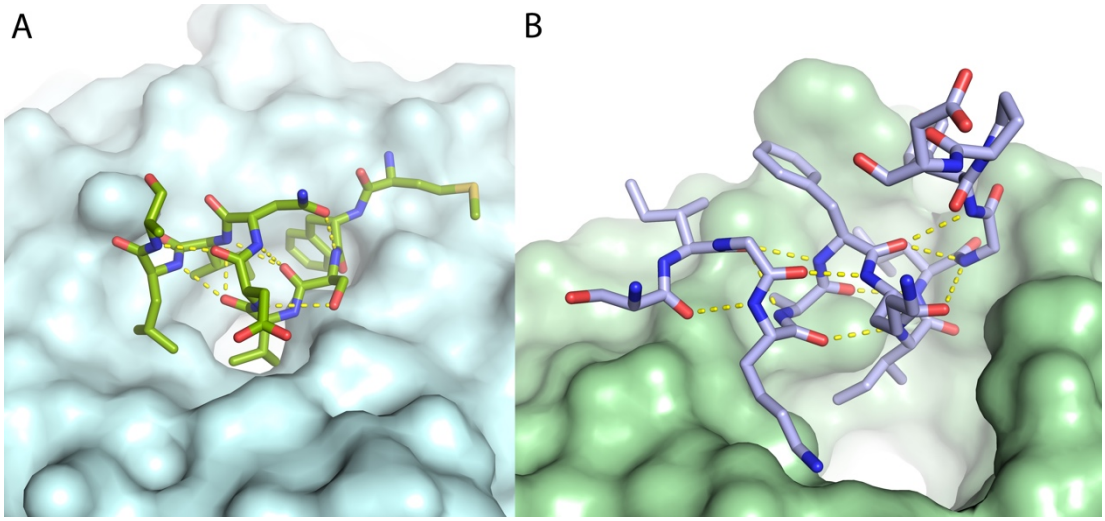

Figure 7.3.3. Binding of EH1 peptide to TLE1 (PDB ID: 2YBA). B) Binding of SIGK peptide to Gβγ (PDB ID: 1XHM). Intramolecular hydrogen bonds are shown as yellow dashed lines.

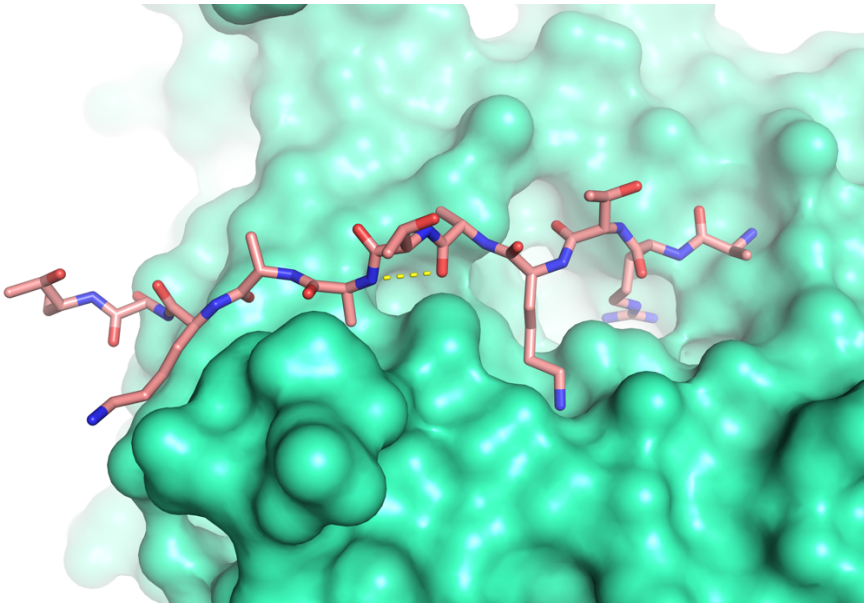

Figure 7.3.4. Binding of histone H3 to Nurf55. Intramolecular hydrogen bonds are shown as yellow dashed lines.

The binding affinity of MeCP2 to TBLR1 (10-20 μM) is lower than other examples of peptide-WD40 interactions. For example, WDR5, discussed earlier in the introduction section, binds MLL1 with an apparent affinity of around 150 nM (Patel et al., 2008), while TolB binds to colicin E9 peptide with an affinity of about 1 μM (Loftus et al., 2006). Nurf55, also shown before, binds histone H3 with affinities between 700 nM and 7 μM, depending on the methylation state of the histone (Nowak et al., 2011). However, there is evidence that some peptides utilise multiple similar binding sites to enhance binding to WD40 domains, for example a low micromolar affinity (10-100 μM) of Sic1 towards its partner Cdc4 can be

overcome by phosphorylation of the peptide at multiple sites, at which point the affinity increases to around 1 µM (Csizmok et al., 2017). It is possible that MeCP2 contains motifs that bind to an allosteric site on the TBLR1. There is a sequence present in full-length MeCP2, dubbed the second NID (SNID) which resembles the NID (amino acids PIKKR in the NID vs IPKKR in the SNID) that could enhance the binding of full-length MeCP2 either to an individual TBLR1 WD40 domain, or to an adjacent WD40 domain in the context of a complete NCoR/SMRT complex with a tetrameric TBLR1 scaffold. It is therefore possible that *in vivo* this interaction is stronger than measured using isolated proteins.

7.3.3 TBLR1 WD40 domain may interact with other proteins

Just as multiple regions of MeCP2 may interact with TBLR1, it is possible that the TBLR1 WD40 domain interacts with other proteins. A large area of TBLR1 circumference shows very high conservation levels, which is an indication of a structural feature. However, there does not appear to be an obvious binding pocket as the side is relatively flat (Figure 7.3.5A). There does exist an electronegative groove that could accommodate a basic peptide similar to MeCP2. The more variable region left of this electronegative area is another potential binding site (Figure 7.3.5A and B).

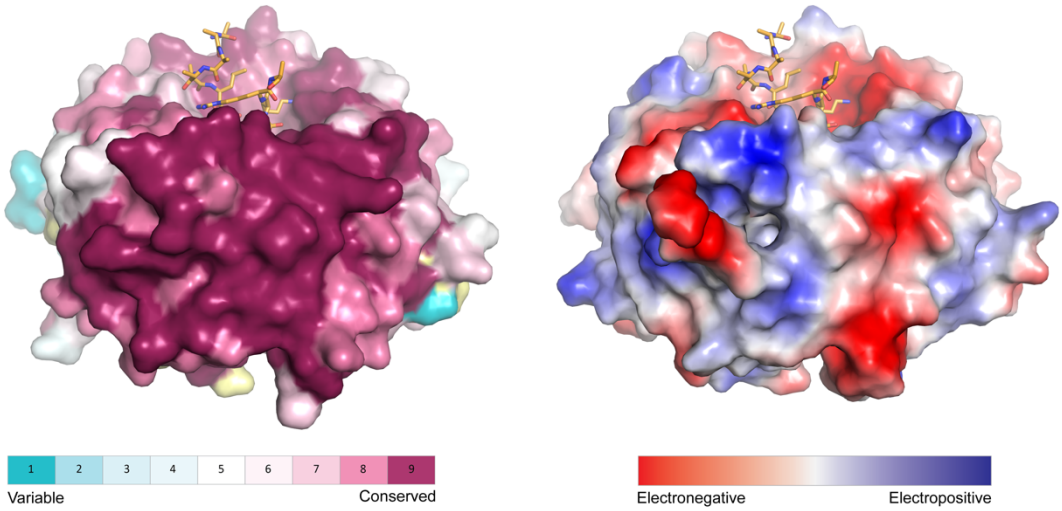

Figure 7.3.5. A) Conservation of mouse TBLR1 WD40 domain. B) Vacuum electrostatics of TBLR1 WD40 domain calculated in PyMol. MeCP2 peptide has been shown in orange.

When peptides do bind to the circumference of a WD40 domain, the binding pocket already exists on the unliganded WD40 domains (Figure 7.3.6). This means that binding of the peptide does not induce a large conformational change in the WD40 domains. In addition, the peptides utilise different strategies for binding, similar to the way peptides bind to the top

face. Even the location of the binding site around the circumference is different between peptides. It is possible that there are proteins or peptides that bind to this area in flat configuration, much like the histone H3 shown above. Consequently, this makes identifying an interaction surface for a potential TBLR1 partner based on the crystal structure alone very difficult.

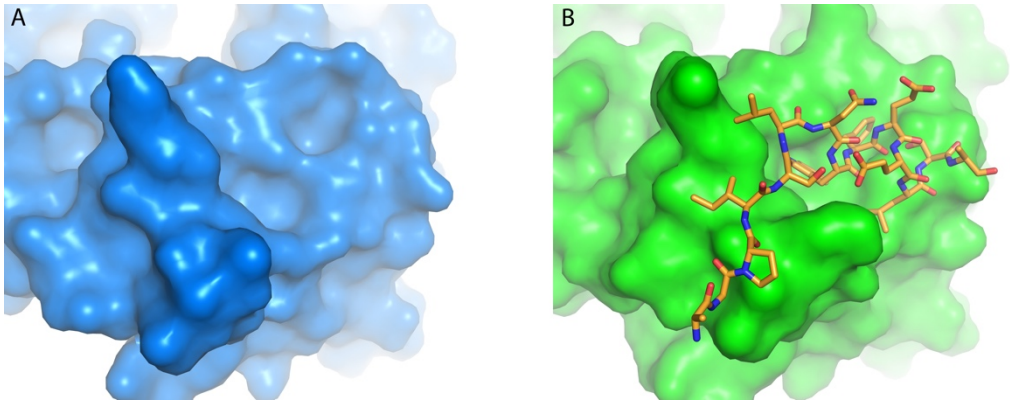

Figure 7.3.6. Comparison of A) unliganded PALB2 (blue) with B) PALB2 (green) bound to BRCA2 peptide (orange).

It is, however, possible to find potential interacting partners that bind to the top face using the NID sequence in a motif search (MOTIF2). One can construct a motif that is likely to bind to the TBLR1 in place of the NID, taking into account the biochemical characteristics of the peptide, namely having a small hydrophobic amino acid after the proline 302, as well as not having a negatively charged amino acid in position 308, as we know that phosphorylation of tyrosine 308 in the wild-type protein abolishes binding. In this case, the motif (in PROSITE pattern format) for searching was “P-[VALI]-K-K-R-x-{DE}.”.

Table 7.3.1. Examples of proteins found by MOTIF2 that contain a motif similar to MeCP2 NID. Protein roles are described as according to UniProt, prediction about motif disorder is based on DisEmbl algorithm.

Protein name	Location	Disordered	Role
PWWP domain-containing protein MUM1	Nucleus	Likely	Nucleosome-binding, DNA-repair
Histone-lysine N-methyltransferase SUV39H2	Nucleus	Yes	Histone lysine methylation
Hepatoma-derived growth factor-related protein 2	Nucleus	Yes	Chromatin binding, DNA binding, positive regulation of cell growth

Transcriptional repressor CTCF	Nucleus	Yes	Transcriptional regulator, nucleosome positioning, regulation of histone methylation and acetylation
Histone-lysine N-methyltransferase ASH1L	Nucleus	Yes	Histone lysine methylation
Nuclear apoptosis-inducing factor 1	Nucleus	Yes	Negative regulator of cell growth

It is intriguing that not only does the search find proteins that are mostly nuclear, many of them are involved in transcriptional regulation and histone modification (Table 7.3.1). In addition, these motifs tend to be positioned in intrinsically disordered regions, a factor that is likely to be required for binding to TBLR1. Given the role of TBLR1 in cells, it is possible that there are other low-affinity binders that compete for binding with MeCP2 to carefully regulate transcription.

7.4 The role of MeCP2 in recruiting TBL1X/TBL1XR1 was investigated

While MeCP2 is a highly conserved protein among vertebrates, the majority of the RTT causing mutations cluster into two domains in the MeCP2 protein – methyl-CpG-binding domain (MBD), and NCoR/SMRT Interaction Domain (NID). The roles of regions beyond these domains were investigated in collaboration with the Bird lab. Rebekah Tillotson from the Bird lab generated knock-in mice that lacked the conserved N-terminus and C-terminus, as well as the region between the MBD and NID. Surprisingly these mice displayed only a very weak RTT phenotype. However, mice that contained both C- and N-terminal truncations displayed no phenotype. This led me to investigate whether the MeCP2 minigene protein was functional, and whether the binding of MeCP2 to TBL1X/TBL1XR1 is cooperative. The analysis of the affinity of MeCP2 for methylated DNA with and without TBL1XR1 was not finished due to time constraints. Despite this, initial results showed that the MeCP2 minigene has some tertiary structure, the functions of MBD and NID are not perturbed, and that the reason why minigene protein carrying mice display RTT phenotype is due to an unknown reason, possibly elevated protein instability as shown by Rebekah Tillotson.

8 Future directions and concluding remarks

8.1 Screening for small-drug like compounds to inhibit or facilitate MeCP2-TBLR1 interaction

Data from experiments done in mice suggests that RTT may be treatable in humans, if the mutated protein could be made to bind TBL1 or TBLR1 more tightly. Similarly, MeCP2 overexpression syndrome should be amenable to treatment by disrupting MeCP2-TBL1/TBLR1 interaction. In either case, small drug-like molecules represent a potential avenue to explore owing to the abundance of drug-like molecule libraries.

MeCP2-TBLR1 interaction can in principle be inhibited in two ways – either directly competing for the MeCP2 binding site on the TBLR1 molecule, or by allosterically modifying the binding site to make MeCP2 unable to bind. As of writing there are only two examples of small drug-like molecule WD40 antagonists in the literature. First is the example of inhibition of WDR5 binding to MLL (Senisterra et al., 2013), where an inhibitor binds to the WDR5 WD40 domain central channel through a combination of hydrophobic interactions and water-mediated hydrogen bonds. The second example is allosteric inhibition of Cdc4 by a small molecule which intercalates between the blades of the WD40 fold and disrupts the substrate binding pocket (Orlicky et al., 2010). In either case, this approach could lead to the treatment of MeCP2 duplication syndrome.

The purpose of facilitating MeCP2-TBLR1 interaction would be to make MeCP2 containing RTT mutations bind to the TBLR1 WD40 domain and therefore treat RTT. For example, a small drug-like molecule could bind in the space left by the R306C RTT mutation. A compound binding in this space should be able to provide the lost hydrogen bonds caused by the arginine-to-cysteine mutation, and one way to achieve this is to have the compound mimic the lost guanidium group of the arginine. It should also be possible to exploit the reactive sulfhydryl group of the R306C mutation to create a covalent linkage between the drug and MeCP2 peptide. An advantage of covalent linkage is a much longer period of action even if the half-life of the compound is short. This also means that a lower or less frequent dose of drug is required for the same effect (Bauer, 2015). Alternatively, these drug-like molecules could facilitate the binding of other RTT mutations by facilitating new binding modes (for P302R) or shielding/neutralising charges that prevent binding (for K304E). In addition,

allosteric binders could alter the central channel topology which could facilitate the binding of other RTT mutations, such as K305R.

8.1.1 Choice of assay

On a very basic level an assay that is used to screen for the binding of drug-like molecules needs to produce a signal (or an absence of a signal) whenever such interaction occurs between the compound and its target. A variety of biochemical assays fulfil this criterion. However, for the purpose of drug development a number of constraints are introduced, such as scalability, speed of setting up, and fast readouts.

Of the assays used during this project, both fluorescence anisotropy and thermal denaturation assays could be used to screen these compounds. In the case of former, binding of a fluorescent MeCP2 peptide to TBLR1 WD40 domain in the presence of increasing concentrations of a compound would reveal whether it could compete MeCP2 off TBLR1 and therefore act as an inhibitor. Alternatively, competing fluorescently labelled MeCP2 peptide off with a non-fluorescent peptide in the presence of compounds would show whether a given compound is able to strengthen the interaction. However, in its current state the fluorescence anisotropy assay I used is not a good choice for high-throughput screening of compounds due to two reasons. First, the interaction between MeCP2 and TBLR1 is rather weak (between 10 and 20 μM), which means a large excess of TBLR1 is required to reach high levels of MeCP2 binding. In addition, currently the assay is done in 96-well plates with well volumes of 100 μl. This combined with the large number of compounds present in chemical libraries (>10^5) makes it currently unfeasible. One way around this would be to scale this assay down to 1536-well plate format with small well volumes (<5 μl) and automatic pipetting using robots. This would both reduce the amount of materials needed and allow to screen a larger number of compounds in the same timeframe while not changing the reliability of the assay (Turconi et al., 2001). The other assay utilised extensively during project, thermal denaturation assay, requires less protein owing to both lower concentration and volume requirement (2 μM, 50 μl). Even then, screening of a large number of compounds is time-consuming unless robotics can be utilised to do the pipetting. A drawback of this assay is that the (de)stabilisation of the target is an indirect readout for binding, and accurate determination of K_D values is difficult (Lo et al., 2004). Therefore, this assay is more suitable for a binary readout (binding/no binding), rather than for the accurate determination of binding affinities.

AlphaLISA is a high-throughput assay that could also be implemented for compound screening. The principle of the assay is to bring two beads (donor and acceptor) coated in antibodies against the protein(s) of interest into proximity. Illuminating the donor beads with a wavelength of 680 nm produces a series of chemical reactions that results in chemiluminescence (Figure 8.1.1). Importantly, this only happens when the two beads are less than 200 nm apart.

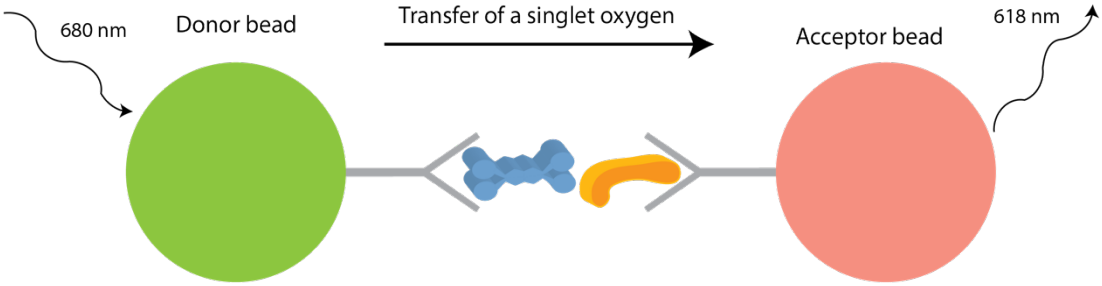

Figure 8.1.1. Principle of AlphaLISA assay. Two proteins, in our case TBLR1 (blue) and MeCP2 (orange) are bound to acceptor and donor beads via antibodies that have been attached to these beads. If the two beads are less than 200 nm apart, i.e. a binding event has occurred, then illumination of the donor bead with a wavelength of 680nm produces a singlet oxygen which initiates a cascade of chemical reactions in the acceptor bead, leading to chemiluminescence at 618 nm that can be detected.

Variations of this assay exist, for example AlphaScreen which uses a slightly different chemical for light production, and which does not require specific antibodies against the protein of interest. Rather, one of the proteins can be biotinylated and attached to the streptavidin-coated donor beads directly while the acceptor beads can be coated in anti-His or similar antibody. It is also possible to link the proteins directly to beads using reductive amination. In our case, the TBLR1-MeCP2 interaction can be probed by attaching these two proteins to the donor and acceptor beads and detecting chemiluminescence in the presence of compounds. This assay could detect either inhibitors of wild-type MeCP2-TBLR1 or activators of mutant MeCP2-TBLR1 interaction. The advantages of this type of screen are high sensitivity, high signal to background ratio, easy miniaturization, as well as being able to perform this assay on crude cell lysates. In the case of AlphaLISA, a drawback is the requirement for specific antibodies towards the interaction partners, which can be costly.

The main drawback of most in vitro assays in high-throughput screening is the requirement for large amount of pure recombinant protein, which is not always easily achieved. An alternative is to use cell-based assays with a reporter that allows to detect the interaction between MeCP2 and TBLR1. These assays have the benefit of being in an environment the

proteins are normally in, as well as revealing potential issues with the test compounds such as membrane permeability and cytotoxicity.

A commonly used technique is two-hybrid screening where one protein is fused to a domain of a transcription factor that binds to the Upstream Activation Sequence (UAS) of a gene, and the other protein fused to an activation domain(AD) (Figure 8.1.2) (Bruckner et al., 2009). Only when the two proteins interact can the transcription factor domains initiate transcription from the reporter gene, which can be detected. One way this could be used in high-throughput screening is to couple MeCP2 (or TBLR1) to the UAS-binding domain, and TBLR1 (or MeCP2) to the Activation Domain, which drives the expression of luciferase. These cells can then be seeded into 1536-well plates, the compounds added, and the luciferase activity measured using a plate-reader. It is possible to detect both inhibitors and facilitators of this interaction by comparing luciferase activity with added compounds against the cells alone - inhibitors should decrease luciferase activity, while activators should increase it.

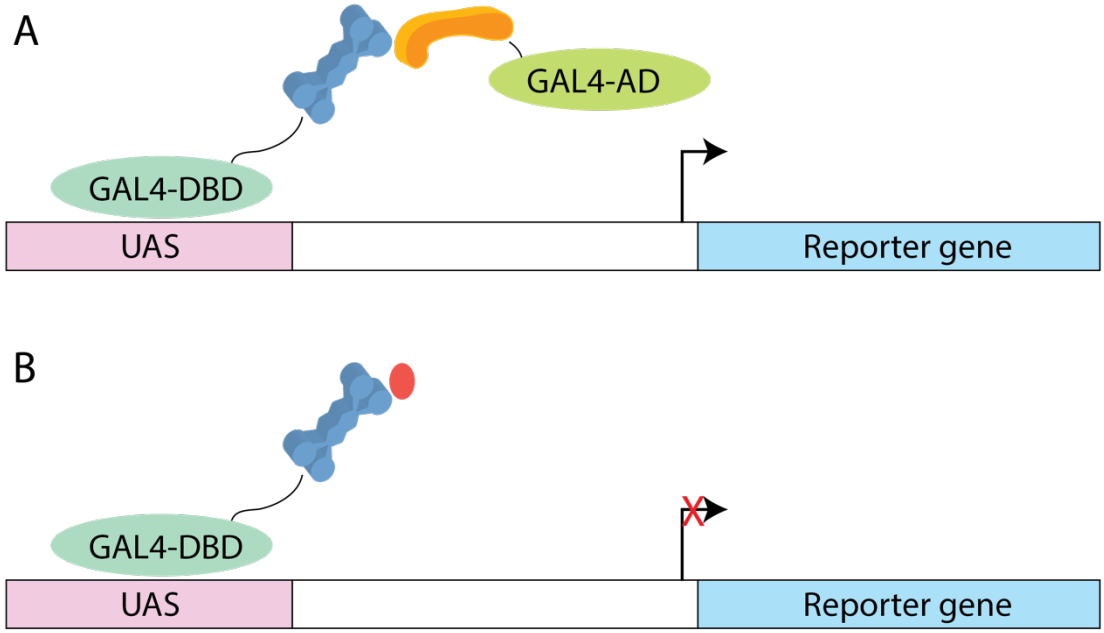

Figure 8.1.2. Example of yeast-two-hybrid assay. A) TBLR1 (blue) is fused to a GAL4-DBD. If MeCP2 (orange), which has been fused to GAL4-AD, interacts with TBLR1, reporter gene is expressed and this can be detected. B) If an inhibitor that binds to TBLR1 blocks MeCP2-GAL4-AD recruitment, reporter gene is not expressed.

Another assay currently in use looks at the recruitment of TBLR1 by MeCP2. This relies on the principle that mCherry-tagged TBL1 co-localises with GFP-tagged MeCP2 in 3T3 cells, but is localised in the cytoplasm upon the loss of MeCP2-TBL1 interaction. By looking at the localisation of TBL1 or TBLR1 in presence of different compounds, it would show which compounds are able to inhibit the co-localisation and therefore the interaction between the wild-type MeCP2 and TBLR1. Alternatively, activators of mutant MeCP2-TBLR1 interaction could similarly be screened.

Finally, *in silico* screens are more and more widely used in high-throughput drug screening as the first step to reduce the number of compounds required to screen, which can reduce the cost of drug development. Due to the availability of the TBLR1 crystal structure with the bound peptide, it is possible to use structure-based virtual screening to define the ligand binding site on TBLR1, and dock chemical compound libraries using specialised software to find potential activators/inhibitors. In addition, it is possible to use the bound MeCP2 as a search model in ligand-based virtual screening, which can be then followed by structure-based approaches. Both approaches have their advantages and limitations, some of which are covered in this review (Scior et al., 2012). However, with careful design it is possible to dramatically reduce the number of compounds needed to test.

8.1.2 Drug-based approach versus current therapeutic approaches

There are currently several different approaches being developed to treat RTT by various groups. One of these approaches is seeking to reintroduce a functional copy of MeCP2 in patients who carry RTT mutations. It has been shown that it is possible to deliver a copy of the MeCP2 gene into null male mice using recombinant adeno-associated virus vectors, which reduces mortality in the null mice while not causing significant defects when introduced into the wild-type mice (Gadalla et al., 2013). A similar strategy on the level of gene therapy is looking to reactivate the silent allele in heterozygous females. However, both of these approaches have the problem associated with dosage, as too much MeCP2 leads to MeCP2 overexpression syndrome which causes symptoms similar to RTT. In addition, as of writing, the reactivation of just a single gene on inactivated X-chromosome is not possible, and the reactivation of a whole X-chromosome will very likely lead to very large gene dosage problems.

On a protein level, another therapeutic approach developed by ArmaGen is to replace MeCP2 in brains by injecting MeCP2 fused to a partner that would allow the protein to cross over the brain-blood barrier (BBB) and reach the neurons to perform its function (ArmaGen).

Difficulties involved in this approach include controlling dosage, as well as ensuring that the MeCP2 retains its function when fused to the partner. Additionally, a very recent new avenue is using RNA-editing to fix MeCP2 mRNA transcripts to make functional protein. However, this research is very much in its early stages and so nothing is known about its efficacy or potential side-effects.

On a drug level there has been widespread research into using compounds that are able to induce read-through of stop codons to allow the normal transcriptional machinery to make functional protein. This could be helpful for many cases of RTT that are caused by nonsense mutations such as R270X and G273X. An issue with these therapies is that they usually use aminoglycosides which are highly toxic, although newer, less-toxic alternatives are being developed (Welch et al., 2007). Furthermore, their efficacies are dependent on the underlying mutations. Therefore, while the concept has been shown to work, it is not a long term solution for these patients.

While these approaches have problems, it is not to say that drug-based approach would be necessarily easier. The main issue with drugs is their permeability to the blood-brain barrier, which is critical for a central nervous system disorder like RTT. In addition, the drugs must have limited side-effects, ideally be orally bioavailable, and most importantly have appropriate kinetics so that the NCoR/SMRT complex is still able to respond to neuronal stimuli and not lock it in a state of constant activity/inactivity. However, small drug-like molecules can be easily titrated, ameliorating issues with dosage that plague some of these other treatment options. In this light these different approaches should not be viewed as competitive, rather they are orthogonal potential therapies that may be able to provide treatment for the myriad of different MeCP2 mutations that cause RTT.

8.2 Does MeCP2 interact with any other parts of NCoR/SMRT co-repressor complex?

In the crystal structure, only residues 298-307 of MeCP2 were seen, while the peptide used in co-crystallisation spanned residues 285-309. From the SPR data it was seen that $MeCP2_{285-313}$ peptide had a slightly higher affinity for TBL1XR1 WD40 domain than the $MeCP2_{285-309}$ peptide. While the amino acids seen in the crystal structure are likely to be the main binding determinants, it is possible that regions beyond N- and/or C-terminus of $MeCP2_{285-309}$ are involved in interaction with either TBL1X/TBL1XR1 WD40 domain directly, or with N-terminal

parts of TBL1X/TBL1XR1. It is also possible that MeCP2 interacts with other components of the NCoR/SMRT complex, such as NCoR/SMRT or HDAC3. A further question is the stoichiometry of the interaction. As there are four WD40 domains in the corepressor complex, it would be interesting to know whether there are four MeCP2 molecules binding to NCoR/SMRT or one MeCP2 molecule binding across two WD40 domains, and whether this binding is cooperative.

8.3 Can mutations in TBL1X or TBL1XR1 cause RTT?

During this project it was shown that binding of MeCP2 to TBL1XR1 can be abolished by mutating amino acids in TBL1XR1, however no mutations in TBL1X or TBL1XR1 have been so far shown to cause RTT syndrome, though there are mutations in these proteins that have been characterised as RTT-like phenotype. There are reports in the literature that propose that MeCP2 has a basic region (including the residues seen in the crystal structure) that interacts with DNA, and that mutations in the NID cause RTT by abolishing the interaction of MeCP2 with DNA. However, in the light of the co-crystal structure where all four NID RTT mutations are seen making extensive contacts with TBL1XR1, it is more likely that the loss of binding of MeCP2 to TBL1XR1 is causative of RTT in the case of NID mutations. Demonstrating that these novel mutations abolish the interaction with MeCP2 would be the first step to elucidating their role in these diseases. In addition, if mice bearing any of the TBL1XR1 (or TBL1X) mutations that cause the loss of binding of MeCP2 were to show RTT phenotype, it would be a very convincing proof for the hypothesis that the loss of MeCP2-TBL1XR1 interaction is the reason why NID mutations cause RTT.

8.4 Do TBL1X and TBL1XR1 WD40 domains have other binding partners?

WD40 domains are common scaffolds for protein-protein interactions through their top, bottom and side faces. It is entirely possible that the interaction with MeCP2 is not the only function of these WD40 domains. Indeed, there are patients diagnosed with a RTT-like phenotype that harbour mutations in TBL1XR1. A girl with West syndrome was shown to have a G70D mutation, while a patient with Pierpont syndrome had a Y446C mutation. An L282P mutation has been found in a patient with autism-spectrum disorder. As so many of these diseases share a common phenotype, it would be interesting to see whether the reason

for the symptoms is the perturbation of MeCP2 function through abolishing the function of TBL1XR1 or NCoR/SMRT complex as a whole, or if there is some unknown factor involved.

8.5 Concluding remarks

Since it was first discovered, MeCP2 has been proposed to have a multifaceted role in cellular context. However, RTT-causing mutations have mostly been only found in two domains of MeCP2. The structure of MeCP2 bound to DNA provided first insight into how MBD mutations cause RTT. Here we have additionally shown how NID mutations can cause RTT. In the process of answering how MeCP2 interacts with NCoR/SMRT corepressor complex, several other fascinating questions about RTT were raised. In the future, some of these will hopefully be answered, leading to a more complete understanding of the protein-protein and protein-DNA interactions involved in RTT and other autism-spectrum disorders.

9 References

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

ADAMS, V. H., MCBRYANT, S. J., WADE, P. A., WOODCOCK, C. L. & HANSEN, J. C. 2007. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J Biol Chem,* 282**,** 15057-64.

AMIR, R. E., VAN DEN VEYVER, I. B., WAN, M., TRAN, C. Q., FRANCKE, U. & ZOGHBI, H. Y. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet,* 23**,** 185-8.

ARMAGEN. Available: http://armagen.com/ [Accessed 13/01/2017].

ARMSTRONG, D., DUNN, J. K., ANTALFFY, B. & TRIVEDI, R. 1995. Selective dendritic alterations in the cortex of Rett syndrome. *J Neuropathol Exp Neurol,* 54**,** 195-201.

BAKER, P. R. & CLAUSER, K. R. *ProteinProspector* [Online]. Available: http://prospector.ucsf.edu/ [Accessed 11/09/2016].

BAKER, S. A., CHEN, L., WILKINS, A. D., YU, P., LICHTARGE, O. & ZOGHBI, H. Y. 2013. An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell,* 152**,** 984-96.

BALLESTAR, E., ROPERO, S., ALAMINOS, M., ARMSTRONG, J., SETIEN, F., AGRELO, R., FRAGA, M. F., HERRANZ, M., AVILA, S., PINEDA, M., MONROS, E. & ESTELLER, M. 2005. The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. *Hum Genet,* 116**,** 91-104.

BANKS, G. C., MOHR, B. & REEVES, R. 1999. The HMG-I(Y) A.T-hook peptide motif confers DNA-binding specificity to a structured chimeric protein. *J Biol Chem,* 274**,** 16536-44.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

BATTYE, T. G., KONTOGIANNIS, L., JOHNSON, O., POWELL, H. R. & LESLIE, A. G. 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr,* 67**,** 271-81.

BAUER, R. A. 2015. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov Today,* 20**,** 1061-73.

BOGDANOVIC, O. & VEENSTRA, G. J. 2009. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma,* 118**,** 549-65.

BROWN, K., SELFRIDGE, J., LAGGER, S., CONNELLY, J., DE SOUSA, D., KERR, A., WEBB, S., GUY, J., MERUSI, C., KOERNER, M. V. & BIRD, A. 2016. The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum Mol Genet,* 25**,** 558-70.

BRUCKNER, A., POLGE, C., LENTZE, N., AUERBACH, D. & SCHLATTNER, U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci,* 10**,** 2763-88.

CHAHROUR, M., JUNG, S. Y., SHAW, C., ZHOU, X., WONG, S. T., QIN, J. & ZOGHBI, H. Y. 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science,* 320**,** 1224-9.

CHAUDHURI, I., SODING, J. & LUPAS, A. N. 2008. Evolution of the beta-propeller fold. *Proteins,* 71**,** 795-803.

CHEN, C. K., CHAN, N. L. & WANG, A. H. 2011. The many blades of the beta-propeller proteins: conserved but versatile. *Trends Biochem Sci,* 36**,** 553-61.

COUVERT, P., BIENVENU, T., AQUAVIVA, C., POIRIER, K., MORAINE, C., GENDROT, C., VERLOES, A., ANDRES, C., LE FEVRE, A. C., SOUVILLE, I., STEFFANN, J., DES PORTES, V., ROPERS, H. H., YNTEMA, H. G., FRYNS, J. P., BRIAULT, S., CHELLY, J. & CHERIF, B. 2001. MECP2 is highly mutated in X-linked mental retardation. *Hum Mol Genet,* 10**,** 941-6.

CSIZMOK, V., ORLICKY, S., CHENG, J., SONG, J., BAH, A., DELGOSHAIE, N., LIN, H., MITTAG, T., SICHERI, F., CHAN, H. S., TYERS, M. & FORMAN-KAY, J. D. 2017. An allosteric conduit facilitates dynamic multisite substrate recognition by the SCFCdc4 ubiquitin ligase. *Nat Commun,* 8**,** 13943.

DAITOKU, H., SAKAMAKI, J. & FUKAMIZU, A. 2011. Regulation of FoxO transcription factors by acetylation and protein-protein interactions. *Biochim Biophys Acta,* 1813**,** 1954-60.

DANIELS, G., LI, Y., GELLERT, L. L., ZHOU, A., MELAMED, J., WU, X., ZHANG, X., ZHANG, D., MERUELO, D., LOGAN, S. K., BASCH, R. & LEE, P. 2014. TBLR1 as an androgen receptor (AR) coactivator selectively activates AR target genes to inhibit prostate cancer growth. *Endocr Relat Cancer,* 21**,** 127-42.

DE RUIJTER, A. J., VAN GENNIP, A. H., CARON, H. N., KEMP, S. & VAN KUILENBURG, A. B. 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J,* 370**,** 737-49.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DECIPHERING DEVELOPMENTAL DISORDER. Available: http://www.ddduk.org/ [Accessed 01/09/2016].

DELCUVE, G. P., KHAN, D. H. & DAVIE, J. R. 2012. Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. *Clin Epigenetics,* 4**,** 5.

DRAGICH, J., HOUWINK-MANVILLE, I. & SCHANEN, C. 2000. Rett syndrome: a surprising result of mutation in MECP2. *Hum Mol Genet,* 9**,** 2365-75.

EBERT, D. H., GABEL, H. W., ROBINSON, N. D., KASTAN, N. R., HU, L. S., COHEN, S., NAVARRO, A. J., LYST, M. J., EKIERT, R., BIRD, A. P. & GREENBERG, M. E. 2013. Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature,* 499**,** 341-5.

EMES, R. D. & PONTING, C. P. 2001. A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration. *Hum Mol Genet,* 10**,** 2813-20.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

FISCHLE, W., TSENG, B. S., DORMANN, H. L., UEBERHEIDE, B. M., GARCIA, B. A., SHABANOWITZ, J., HUNT, D. F., FUNABIKI, H. & ALLIS, C. D. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature,* 438**,** 1116-22.

FRAGA, M. F., BALLESTAR, E., MONTOYA, G., TAYSAVANG, P., WADE, P. A. & ESTELLER, M. 2003. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res,* 31**,** 1765-74.

GABEL, H. W., KINDE, B., STROUD, H., GILBERT, C. S., HARMIN, D. A., KASTAN, N. R., HEMBERG, M., EBERT, D. H. & GREENBERG, M. E. 2015. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature,* 522**,** 89-93.

GADALLA, K. K., BAILEY, M. E., SPIKE, R. C., ROSS, P. D., WOODARD, K. T., KALBURGI, S. N., BACHABOINA, L., DENG, J. V., WEST, A. E., SAMULSKI, R. J., GRAY, S. J. & COBB, S. R. 2013. Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice. *Mol Ther,* 21**,** 18-30.

GARCIA-HIGUERA, I., GAITATZES, C., SMITH, T. F. & NEER, E. J. 1998. Folding a WD repeat propeller. Role of highly conserved aspartic acid residues in the G protein beta subunit and Sec13. *J Biol Chem,* 273**,** 9041-9.

GARG, A. & AGGARWAL, B. B. 2002. Nuclear transcription factor-kappaB as a target for cancer drug development. *Leukemia,* 16**,** 1053-68.

GAYATRI, S. & BEDFORD, M. T. 2014. Readers of histone methylarginine marks. *Biochim Biophys Acta,* 1839**,** 702-10.

GHOSH, R. P., HOROWITZ-SCHERER, R. A., NIKITINA, T., GIERASCH, L. M. & WOODCOCK, C. L. 2008. Rett Syndrome-causing Mutations in Human MeCP2 Result in Diverse Structural Changes That Impact Folding and DNA Interactions*. *J Biol Chem,* 283**,** 20523-34.

GUY, J., GAN, J., SELFRIDGE, J., COBB, S. & BIRD, A. 2007. Reversal of neurological defects in a mouse model of Rett syndrome. *Science,* 315**,** 1143-7.

HAN, Z., XING, X., HU, M., ZHANG, Y., LIU, P. & CHAI, J. 2007. Structural basis of EZH2 recognition by EED. *Structure,* 15**,** 1306-15.

HECKMAN, L. D., CHAHROUR, M. H. & ZOGHBI, H. Y. 2014. Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. *Elife,* 3.

HEINEN, C. A., JONGEJAN, A., WATSON, P. J., REDEKER, B., BOELEN, A., BOUDZOVITCH-SUROVTSEVA, O., FORZANO, F., HORDIJK, R., KELLEY, R., OLNEY, A. H., PIERPONT, M. E., SCHAEFER, G. B., STEWART, F., VAN TROTSENBURG, A. S., FLIERS, E., SCHWABE, J. W. & HENNEKAM, R. C. 2016. A

specific mutation in TBL1XR1 causes Pierpont syndrome. *J Med Genet,* 53**,** 330-7.

HERMANN, A., GOYAL, R. & JELTSCH, A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem,* 279**,** 48350-9.

HO, K. L., MCNAE, I. W., SCHMIEDEBERG, L., KLOSE, R. J., BIRD, A. P. & WALKINSHAW, M. D. 2008. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell,* 29**,** 525-31.

HONG, L., SCHROTH, G. P., MATTHEWS, H. R., YAU, P. & BRADBURY, E. M. 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem,* 268**,** 305-14.

HUANG, J., CARDAMONE, M. D., JOHNSON, H. E., NEAULT, M., CHAN, M., FLOYD, Z. E., MALLETTE, F. A. & PERISSI, V. 2015. Exchange Factor TBL1 and Arginine Methyltransferase PRMT6 Cooperate in Protecting G Protein Pathway Suppressor 2 (GPS2) from Proteasomal Degradation. *J Biol Chem,* 290**,** 19044-54.

ILLINGWORTH, R., KERR, A., DESOUSA, D., JORGENSEN, H., ELLIS, P., STALKER, J., JACKSON, D., CLEE, C., PLUMB, R., ROGERS, J., HUMPHRAY, S., COX, T., LANGFORD, C. & BIRD, A. 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol,* 6**,** e22.

JACOBSON, R. H., LADURNER, A. G., KING, D. S. & TJIAN, R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science,* 288**,** 1422-5.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet,* 33 Suppl**,** 245-54.

KABSCH, W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

KANTARDJIEFF, K. A. & RUPP, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci,* 12**,** 1865-71.

KINDE, B., GABEL, H. W., GILBERT, C. S., GRIFFITH, E. C. & GREENBERG, M. E. 2015. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A,* 112**,** 6800-6.

KLOSE, R. J. & BIRD, A. P. 2004. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem,* 279**,** 46490-6.

KLOSE, R. J., SARRAF, S. A., SCHMIEDEBERG, L., MCDERMOTT, S. M., STANCHEVA, I. & BIRD, A. P. 2005. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell,* 19**,** 667-78.

LAKOWICZ, J. R. 2006. *Principles of Fluorescence Spectroscopy*, Springer.

LAVOIE, H., DEBEANE, F., TRINH, Q. D., TURCOTTE, J. F., CORBEIL-GIRARD, L. P., DICAIRE, M. J., SAINT-DENIS, A., PAGE, M., ROULEAU, G. A. & BRAIS, B. 2003.

Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. *Hum Mol Genet,* 12**,** 2967-79.

LEFSTIN, J. A. & YAMAMOTO, K. R. 1998. Allosteric effects of DNA on transcriptional regulators. *Nature,* 392**,** 885-8.

LEWIS, J. D., MEEHAN, R. R., HENZEL, W. J., MAURER-FOGY, I., JEPPESEN, P., KLEIN, F. & BIRD, A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell,* 69**,** 905-14.

LO, M. C., AULABAUGH, A., JIN, G., COWLING, R., BARD, J., MALAMAS, M. & ELLESTAD, G. 2004. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem,* 332**,** 153-9.

LOFTUS, S. R., WALKER, D., MATE, M. J., BONSOR, D. A., JAMES, R., MOORE, G. R. & KLEANTHOUS, C. 2006. Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proc Natl Acad Sci U S A,* 103**,** 12353-8.

LYST, M. J. & BIRD, A. 2015. Rett syndrome: a complex disorder with simple roots. *Nat Rev Genet,* 16**,** 261-75.

LYST, M. J., EKIERT, R., EBERT, D. H., MERUSI, C., NOWAK, J., SELFRIDGE, J., GUY, J., KASTAN, N. R., ROBINSON, N. D., DE LIMA ALVES, F., RAPPSILBER, J., GREENBERG, M. E. & BIRD, A. 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci,* 16**,** 898-902.

MARCHETTO, M. C., CARROMEU, C., ACAB, A., YU, D., YEO, G. W., MU, Y., CHEN, G., GAGE, F. H. & MUOTRI, A. R. 2010. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell,* 143**,** 527-39.

MARGUERON, R., JUSTIN, N., OHNO, K., SHARPE, M. L., SON, J., DRURY, W. J., 3RD, VOIGT, P., MARTIN, S. R., TAYLOR, W. R., DE MARCO, V., PIRROTTA, V., REINBERG, D. & GAMBLIN, S. J. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature,* 461**,** 762-7.

MASCOT. Available: http://www.matrixscience.com/ [Accessed 11/09/2016].

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

MELLEN, M., AYATA, P., DEWELL, S., KRIAUCIONIS, S. & HEINTZ, N. 2012. MeCP2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system. *Cell,* 151**,** 1417-30.

MIGLIORI, V., MULLER, J., PHALKE, S., LOW, D., BEZZI, M., MOK, W. C., SAHU, S. K., GUNARATNE, J., CAPASSO, P., BASSI, C., CECATIELLO, V., DE MARCO, A., BLACKSTOCK, W., KUZNETSOV, V., AMATI, B., MAPELLI, M. & GUCCIONE, E. 2012. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol,* 19**,** 136-44.

MORETTI, P., LEVENSON, J. M., BATTAGLIA, F., ATKINSON, R., TEAGUE, R., ANTALFFY, B., ARMSTRONG, D., ARANCIO, O., SWEATT, J. D. & ZOGHBI, H. Y. 2006. Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J Neurosci,* 26**,** 319-27.

MOTIF2. Available: http://www.genome.jp/tools/motif/MOTIF2.html [Accessed 11/01/2017].

NAN, X., MEEHAN, R. R. & BIRD, A. 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res,* 21**,** 4886-92.

NAN, X., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-9.

NEUL, J. L., FANG, P., BARRISH, J., LANE, J., CAEG, E., SMITH, E. O., ZOGHBI, H., PERCY, A. & GLAZE, D. G. 2008. Specific Mutations in Methyl-CpG-Binding Protein 2 Confer Different Severity in Rett Syndrome. *Neurology,* 70**,** 1313-21.

NEUL, J. L., KAUFMANN, W. E., GLAZE, D. G., CHRISTODOULOU, J., CLARKE, A. J., BAHI-BUISSON, N., LEONARD, H., BAILEY, M. E. S., SCHANEN, N. C., ZAPPELLA, M., RENIERI, A., HUPPKE, P. & PERCY, A. K. 2010. Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. *Ann Neurol,* 68**,** 944-50.

NOTT, A., CHENG, J., GAO, F., LIN, Y. T., GJONESKA, E., KO, T., MINHAS, P., ZAMUDIO, A. V., MENG, J., ZHANG, F., JIN, P. & TSAI, L. H. 2016. Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. *Nat Neurosci*.

NOWAK, A. J., ALFIERI, C., STIRNIMANN, C. U., RYBIN, V., BAUDIN, F., LY-HARTIG, N., LINDNER, D. & MULLER, C. W. 2011. Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. *J Biol Chem,* 286**,** 23388-96.

O'ROAK, B. J., VIVES, L., GIRIRAJAN, S., KARAKOC, E., KRUMM, N., COE, B. P., LEVY, R., KO, A., LEE, C., SMITH, J. D., TURNER, E. H., STANAWAY, I. B., VERNOT, B., MALIG, M., BAKER, C., REILLY, B., AKEY, J. M., BORENSTEIN, E., RIEDER, M. J., NICKERSON, D. A., BERNIER, R., SHENDURE, J. & EICHLER, E. E. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature,* 485**,** 246-50.

OBEROI, J., FAIRALL, L., WATSON, P. J., YANG, J. C., CZIMMERER, Z., KAMPMANN, T., GOULT, B. T., GREENWOOD, J. A., GOOCH, J. T., KALLENBERGER, B. C., NAGY, L., NEUHAUS, D. & SCHWABE, J. W. 2011. Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nat Struct Mol Biol,* 18**,** 177-84.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

ORLICKY, S., TANG, X., NEDUVA, V., ELOWE, N., BROWN, E. D., SICHERI, F. & TYERS, M. 2010. An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase. *Nat Biotechnol,* 28**,** 733-7.

PATEL, A., VOUGHT, V. E., DHARMARAJAN, V. & COSGROVE, M. S. 2008. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem,* 283**,** 32162-75.

PATEL, D. J. & WANG, Z. 2013. Readout of epigenetic modifications. *Annu Rev Biochem,* 82**,** 81-118.

PERISSI, V., AGGARWAL, A., GLASS, C. K., ROSE, D. W. & ROSENFELD, M. G. 2004. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell,* 116**,** 511-26.

PERKINS, N. D. 2004. Regulation of NF-kappaB by atypical activators and tumour suppressors. *Biochem Soc Trans,* 32**,** 936-9.

PONCET-MONTANGE, G., ZHAN, Y., BARDENHAGEN, J. P., PETROCCHI, A., LEO, E., SHI, X., LEE, G. R. T., LEONARD, P. G., GECK DO, M. K., CARDOZO, M. G., ANDERSEN, J. N., PALMER, W. S., JONES, P. & LADBURY, J. E. 2015. Observed bromodomain flexibility reveals histone peptide- and small molecule ligand-compatible forms of ATAD2. *Biochem J,* 466**,** 337-46.

POPE, A. J., HAUPTS, U. M. & MOORE, K. J. 1999. Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. *Drug Discov Today,* 4**,** 350-362.

RAIBER, E. A., MURAT, P., CHIRGADZE, D. Y., BERALDI, D., LUISI, B. F. & BALASUBRAMANIAN, S. 2015. 5-Formylcytosine alters the structure of the DNA double helix. *Nat Struct Mol Biol,* 22**,** 44-9.

RAMADOSS, S., LI, J., DING, X., AL HEZAIMI, K. & WANG, C. Y. 2011. Transducin beta-like protein 1 recruits nuclear factor kappaB to the target gene promoter for transcriptional activation. *Mol Cell Biol,* 31**,** 924-34.

RARE CODON CALCULATOR. Available: http://nihserver.mbi.ucla.edu/RACC/ [Accessed 16/08/2016].

RETTBASE. Available: http://mecp2.chw.edu.au/ [Accessed 07/06/2016].

ROSE, N. R. & KLOSE, R. J. 2014. Understanding the relationship between DNA methylation and histone lysine methylation☆. *Biochim Biophys Acta,* 1839**,** 1362-72.

RUTHENBURG, A. J., WANG, W., GRAYBOSCH, D. M., LI, H., ALLIS, C. D., PATEL, D. J. & VERDINE, G. L. 2006. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol,* 13**,** 704-12.

SAITSU, H., TOHYAMA, J., WALSH, T., KATO, M., KOBAYASHI, Y., LEE, M., TSURUSAKI, Y., MIYAKE, N., GOTO, Y., NISHINO, I., OHTAKE, A., KING, M. C. & MATSUMOTO, N. 2014. A girl with West syndrome and autistic features harboring a de novo TBL1XR1 mutation. *J Hum Genet,* 59**,** 581-3.

SALICHS, E., LEDDA, A., MULARONI, L., ALBA, M. M. & DE LA LUNA, S. 2009. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet,* 5**,** e1000397.

SAMACO, R. C. & NEUL, J. L. 2011. Complexities of Rett syndrome and MeCP2. *J Neurosci,* 31**,** 7951-9.

SCIOR, T., BENDER, A., TRESADERN, G., MEDINA-FRANCO, J. L., MARTINEZ-MAYORGA, K., LANGER, T., CUANALO-CONTRERAS, K. & AGRAFIOTIS, D. K. 2012. Recognizing pitfalls in virtual screening: a critical review. *J Chem Inf Model,* 52**,** 867-81.

SENISTERRA, G., WU, H., ALLALI-HASSANI, A., WASNEY, G. A., BARSYTE-LOVEJOY, D., DOMBROVSKI, L., DONG, A., NGUYEN, K. T., SMIL, D., BOLSHAN, Y., HAJIAN, T., HE, H., SEITOVA, A., CHAU, I., LI, F., PODA, G., COUTURE, J. F., BROWN, P. J., AL-AWAR, R., SCHAPIRA, M., ARROWSMITH, C. H. & VEDADI, M. 2013. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochem J,* 449**,** 151-9.

SHAHBAZIAN, M., YOUNG, J., YUVA-PAYLOR, L., SPENCER, C., ANTALFFY, B., NOEBELS, J., ARMSTRONG, D., PAYLOR, R. & ZOGHBI, H. 2002. Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron,* 35**,** 243-54.

SKENE, P. J., ILLINGWORTH, R. S., WEBB, S., KERR, A., JAMES, K. D., TURNER, D. J., ANDREWS, R. & BIRD, A. P. 2010. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell,* 37**,** 457-68.

SMITH, Z. D. & MEISSNER, A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet,* 14**,** 204-20.

SONG, J. J. & KINGSTON, R. E. 2008. WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem,* 283**,** 35258-64.

STIRNIMANN, C. U., PETSALAKI, E., RUSSELL, R. B. & MULLER, C. W. 2010. WD40 proteins propel cellular networks. *Trends Biochem Sci,* 35**,** 565-74.

SUTER, B., TREADWELL-DEERING, D., ZOGHBI, H. Y., GLAZE, D. G. & NEUL, J. L. 2014. Brief report: MECP2 mutations in people without Rett syndrome. *J Autism Dev Disord,* 44**,** 703-11.

SUZUKI, M. M. & BIRD, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet,* 9**,** 465-76.

SZULWACH, K. E., LI, X., LI, Y., SONG, C. X., WU, H., DAI, Q., IRIER, H., UPADHYAY, A. K., GEARING, M., LEVEY, A. I., VASANTHAKUMAR, A., GODLEY, L. A., CHANG, Q., CHENG, X., HE, C. & JIN, P. 2011. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci,* 14**,** 1607-16.

TEMUDO, T., SANTOS, M., RAMOS, E., DIAS, K., VIEIRA, J. P., MOREIRA, A., CALADO, E., CARRILHO, I., OLIVEIRA, G., LEVY, A., BARBOT, C., FONSECA, M., CABRAL, A., CABRAL, P., MONTEIRO, J., BORGES, L., GOMES, R., MIRA, G., PEREIRA, S. A., SANTOS, M., FERNANDES, A., EPPLEN, J. T., SEQUEIROS, J. & MACIEL, P. 2011. Rett syndrome with and without detected MECP2 mutations: an attempt to redefine phenotypes. *Brain Dev,* 33**,** 69-76.

THE HUMAN PROTEIN ATLAS. [Accessed 10/08/16 2016].

TRAPPE, R., LACCONE, F., COBILANSCHI, J., MEINS, M., HUPPKE, P., HANEFELD, F. & ENGEL, W. 2001. MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin. *Am J Hum Genet,* 68**,** 1093-101.

TURCONI, S., SHEA, K., ASHMAN, S., FANTOM, K., EARNSHAW, D. L., BINGHAM, R. P., HAUPTS, U. M., BROWN, M. J. & POPE, A. J. 2001. Real experiences of uHTS: a prototypic 1536-well fluorescence anisotropy-based uHTS screen and application of well-level quality control procedures. *J Biomol Screen,* 6**,** 275-90.

TURNER, B. M. 2007. Defining an epigenetic code. *Nat Cell Biol,* 9**,** 2-6.

VAN ESCH, H. 2012. MECP2 Duplication Syndrome. *Mol Syndromol,* 2**,** 128-136.

VAN ESCH, H., BAUTERS, M., IGNATIUS, J., JANSEN, M., RAYNAUD, M., HOLLANDERS, K., LUGTENBERG, D., BIENVENU, T., JENSEN, L. R., GECZ, J., MORAINE, C., MARYNEN, P., FRYNS, J. P. & FROYEN, G. 2005. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet,* 77**,** 442-53.

VANDERMARLIERE, E., BOURGOIS, T. M., WINN, M. D., VAN CAMPENHOUT, S., VOLCKAERT, G., DELCOUR, J. A., STRELKOV, S. V., RABIJNS, A. & COURTIN, C. M. 2009. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem J,* 418**,** 39-47.

VILLARD, L. 2007. MECP2 mutations in males. *J Med Genet,* 44**,** 417-23.

VILLARD, L., KPEBE, A., CARDOSO, C., CHELLY, P. J., TARDIEU, P. M. & FONTES, M. 2000. Two affected boys in a Rett syndrome family: clinical and molecular findings. *Neurology,* 55**,** 1188-93.

WAKEFIELD, R. I., SMITH, B. O., NAN, X., FREE, A., SOTERIOU, A., UHRIN, D., BIRD, A. P. & BARLOW, P. N. 1999. The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol,* 291**,** 1055-65.

WALPORT, L. J., HOPKINSON, R. J. & SCHOFIELD, C. J. 2012. Mechanisms of human histone and nucleic acid demethylases. *Curr Opin Chem Biol,* 16**,** 525-34.

WANG, J., OU, J., GUO, Y., DAI, T., LI, X., LIU, J., XIA, M., LIU, L. & HE, M. 2014. TBLR1 is a novel prognostic marker and promotes epithelial-mesenchymal transition in cervical cancer. *Br J Cancer,* 111**,** 112-24.

WANG, Y., HU, X. J., ZOU, X. D., WU, X. H., YE, Z. Q. & WU, Y. D. 2015. WDSPdb: a database for WD40-repeat proteins. *Nucleic Acids Res,* 43**,** D339-44.

WANG, Y., JIANG, F., ZHUO, Z., WU, X. H. & WU, Y. D. 2013. A method for WD40 repeat detection and secondary structure prediction. *PLoS One,* 8**,** e65705.

WATSON, P. J., FAIRALL, L., SANTOS, G. M. & SCHWABE, J. W. 2012. Structure of HDAC3 bound to co-repressor and inositol tetraphosphate. *Nature,* 481**,** 335-40.

WEBER, M., HELLMANN, I., STADLER, M. B., RAMOS, L., PAABO, S., REBHAN, M. & SCHUBELER, D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet,* 39**,** 457-66.

WELCH, E. M., BARTON, E. R., ZHUO, J., TOMIZAWA, Y., FRIESEN, W. J., TRIFILLIS, P., PAUSHKIN, S., PATEL, M., TROTTA, C. R., HWANG, S., WILDE, R. G., KARP, G., TAKASUGI, J., CHEN, G., JONES, S., REN, H., MOON, Y. C., CORSON, D., TURPOFF, A. A., CAMPBELL, J. A., CONN, M. M., KHAN, A., ALMSTEAD, N. G., HEDRICK, J., MOLLIN, A., RISHER, N., WEETALL, M., YEH, S., BRANSTROM, A. A., COLACINO, J. M., BABIAK, J., JU, W. D., HIRAWAT, S., NORTHCUTT, V. J., MILLER, L. L., SPATRICK, P., HE, F., KAWANA, M., FENG, H., JACOBSON, A., PELTZ, S. W. & SWEENEY, H. L. 2007. PTC124 targets genetic disorders caused by nonsense mutations. *Nature,* 447**,** 87-91.

WHITTLE, J. R. & SCHWARTZ, T. U. 2010. Structure of the Sec13-Sec16 edge element, a template for assembly of the COPII vesicle coat. *J Cell Biol,* 190**,** 347-61.

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G. W., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-42.

WU, H. & ZHANG, Y. 2015. Charting oxidized methylcytosines at base resolution. *Nat Struct Mol Biol,* 22**,** 656-61.

WU, X. H., CHEN, R. C., GAO, Y. & WU, Y. D. 2010. The effect of Asp-His-Ser/Thr-Trp tetrad on the thermostability of WD40-repeat proteins. *Biochemistry,* 49**,** 10237-45.

WU, X. H., WANG, Y., ZHUO, Z., JIANG, F. & WU, Y. D. 2012. Identifying the hotspots on the top faces of WD40-repeat proteins from their primary sequences by beta-bulges and DHSW tetrads. *PLoS One,* 7**,** e43005.

XU, C., TEMPEL, W., HE, H., WU, X., BOUNTRA, C., ARROWSMITH, C. H., EDWARDS, A. M. & MIN, J. 2013. Crystal structure of TBL1XR1 WD40 repeats. *Structural Genomics Consortium*.

YAN, H. T., SHINKA, T., KINOSHITA, K., SATO, Y., UMENO, M., CHEN, G., TSUJI, K., UNEMI, Y., YANG, X. J., IWAMOTO, T. & NAKAHORI, Y. 2005. Molecular analysis of TBL1Y, a Y-linked homologue of TBL1X related with X-linked late-onset sensorineural deafness. *J Hum Genet,* 50**,** 175-81.

YOON, H. G., CHOI, Y., COLE, P. A. & WONG, J. 2005. Reading and Function of a Histone Code Involved in Targeting Corepressor Complexes for Repression. *Mol Cell Biol,* 25**,** 324-35.

YOUNG, J. I., HONG, E. P., CASTLE, J. C., CRESPO-BARRETO, J., BOWMAN, A. B., ROSE, M. F., KANG, D., RICHMAN, R., JOHNSON, J. M., BERGET, S. & ZOGHBI, H. Y. 2005. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A,* 102**,** 17551-8.

ZHANG, L., LU, X., LU, J., LIANG, H., DAI, Q., XU, G. L., LUO, C., JIANG, H. & HE, C. 2012. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol,* 8**,** 328-30.

ZOCCHI, L. & SASSONE-CORSI, P. 2012. SIRT1-mediated deacetylation of MeCP2 contributes to BDNF expression. *Epigenetics,* 7**,** 695-700.

10 Appendix

10.1 TBL1XR1 and MeCP2 SPR data

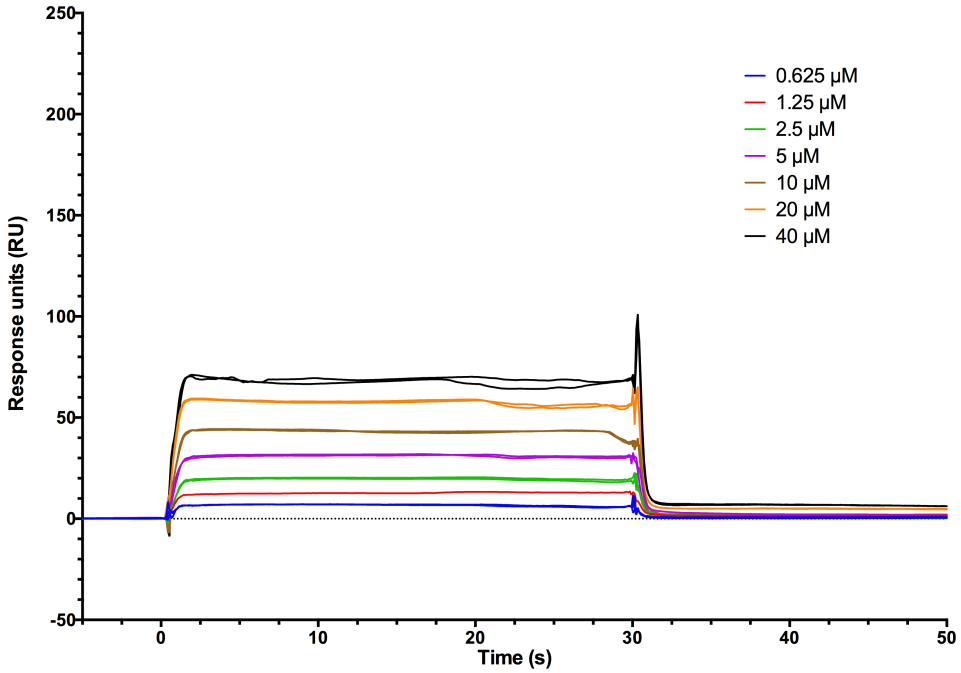

Figure 10.1.1 Wild-type TBL1XR1 with MeCP2 285-313 peptide

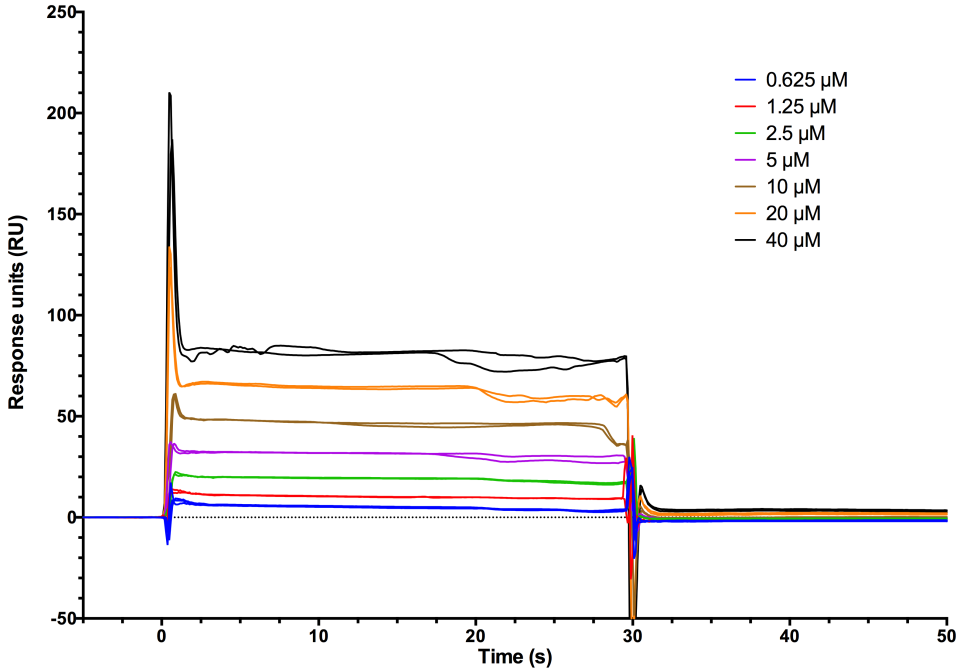

Figure 10.1.2. Wild-type TBL1XR1 with MeCP2 285-309 peptide

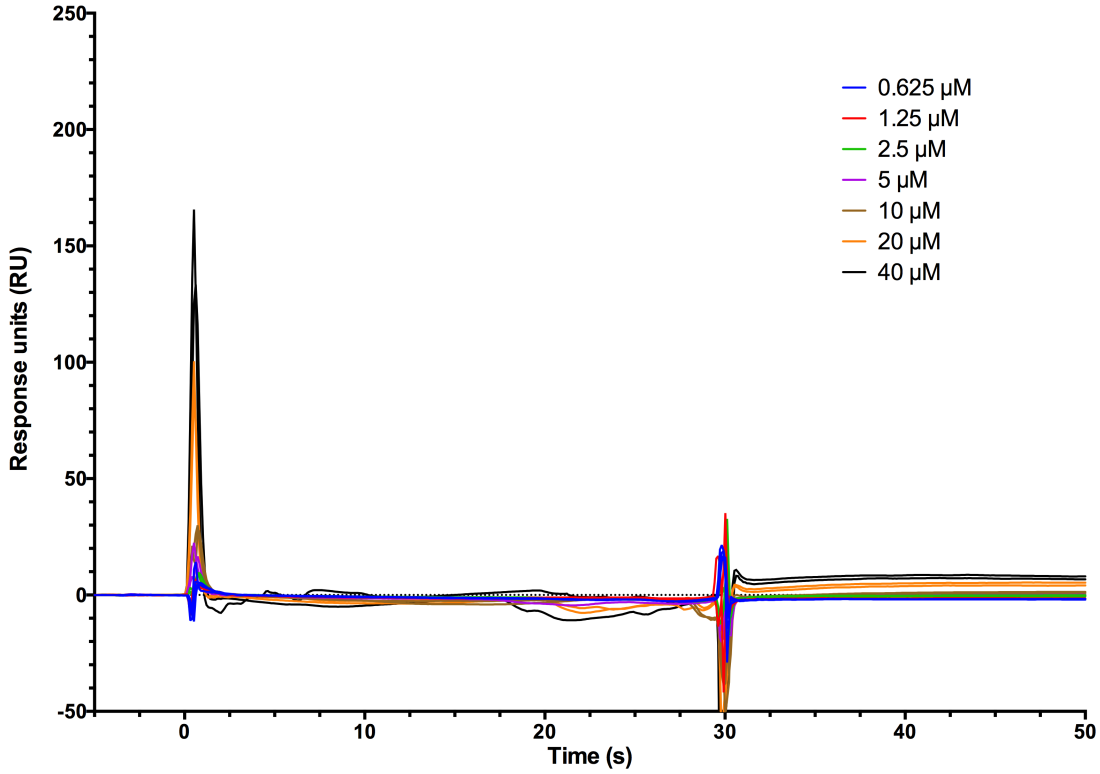

Figure 10.1.3. Wild-type TBL1XR1 with MeCP2 285-313 R306C peptide

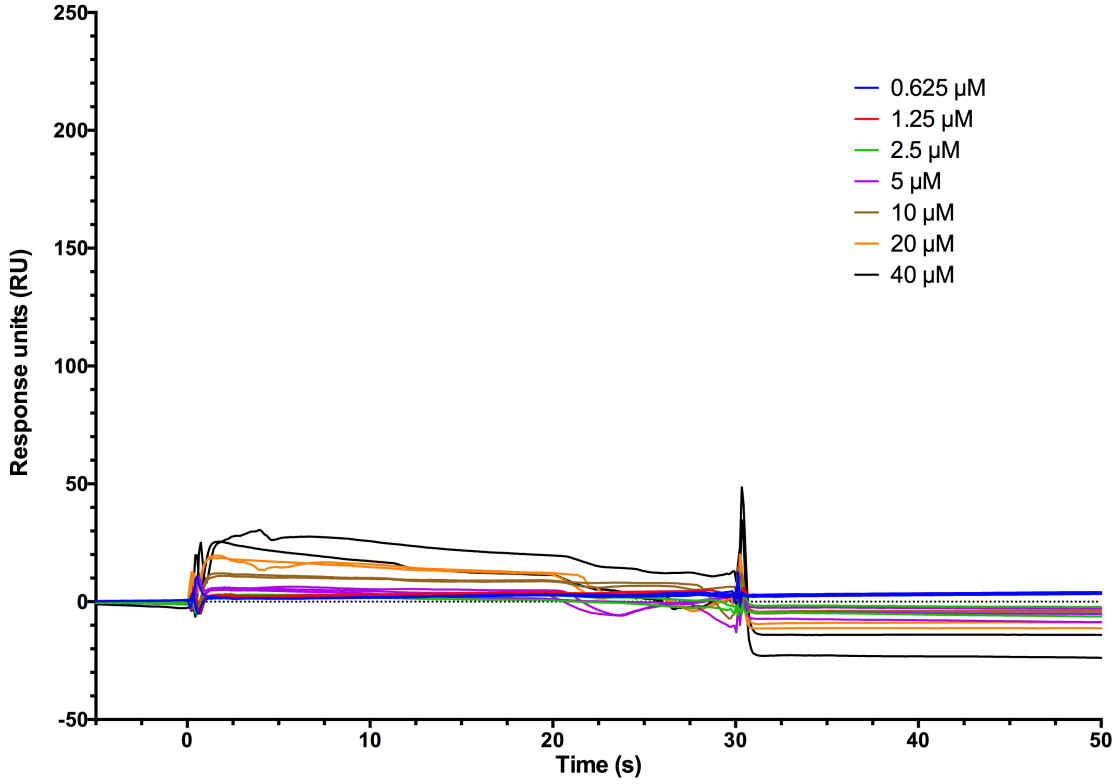

Figure 10.1.4. TBL1XR1 E171A with MeCP2 285-313 peptide

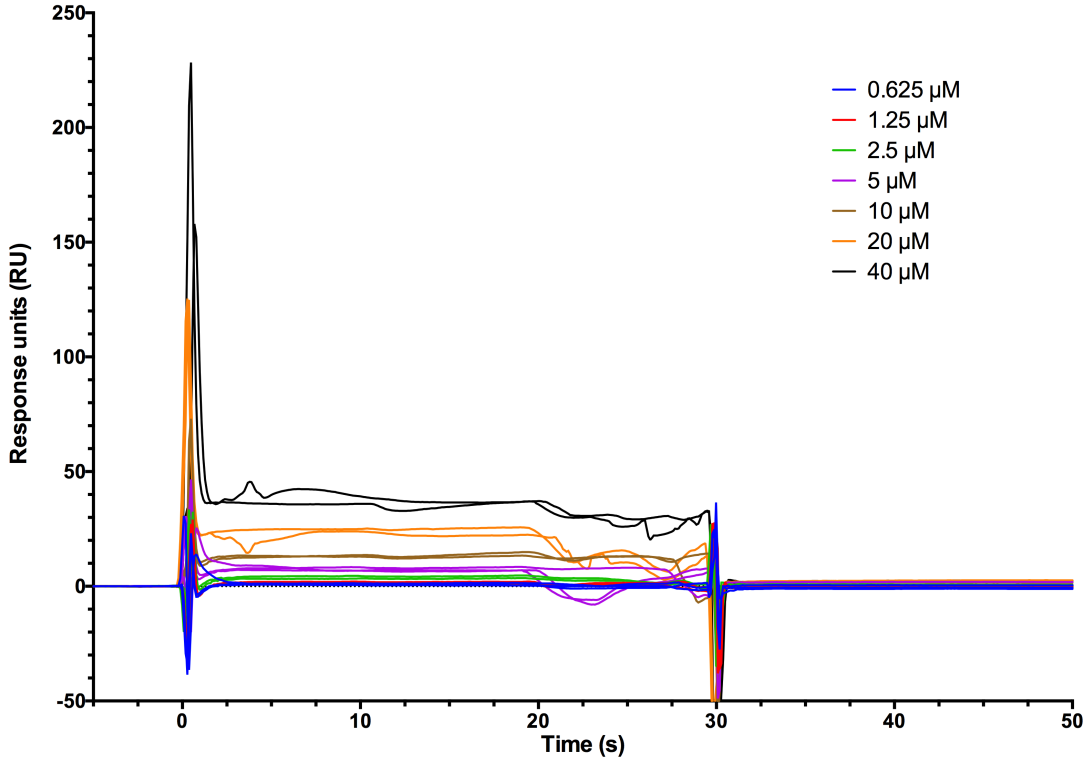

Figure 10.1.5. TBL1XR1 E171A with MeCP2 285-309 peptide

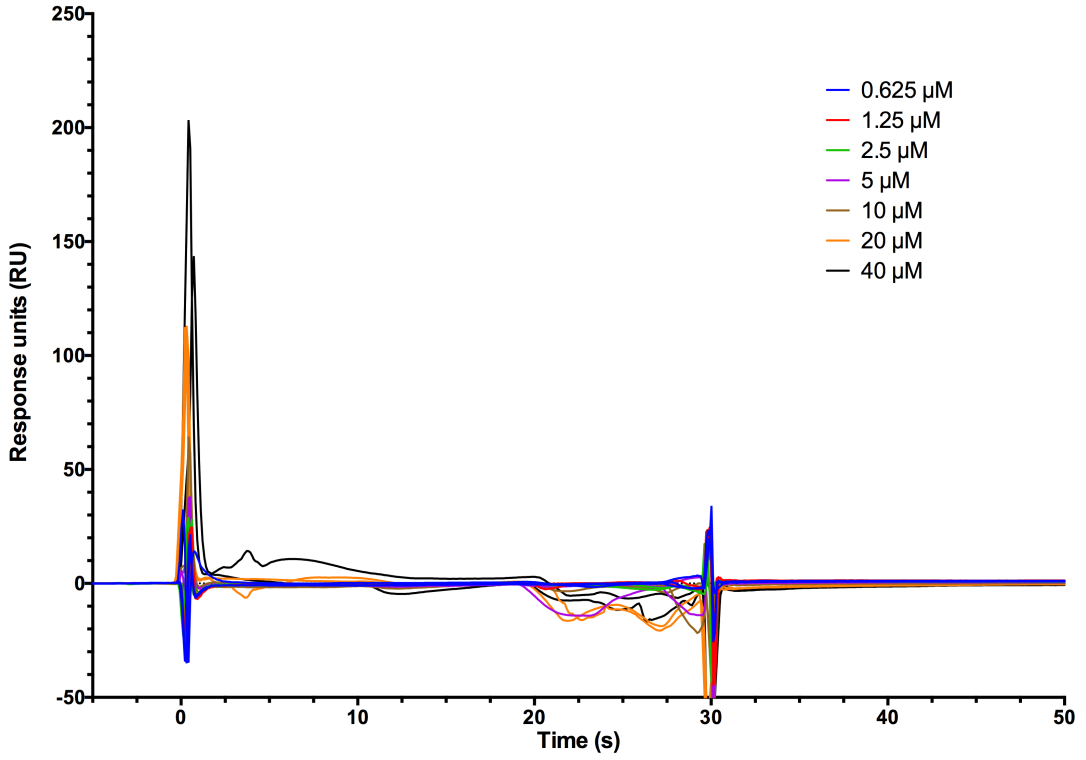

Figure 10.1.6. TBL1XR1 E171A with MeCP2 285-313 R306C peptide

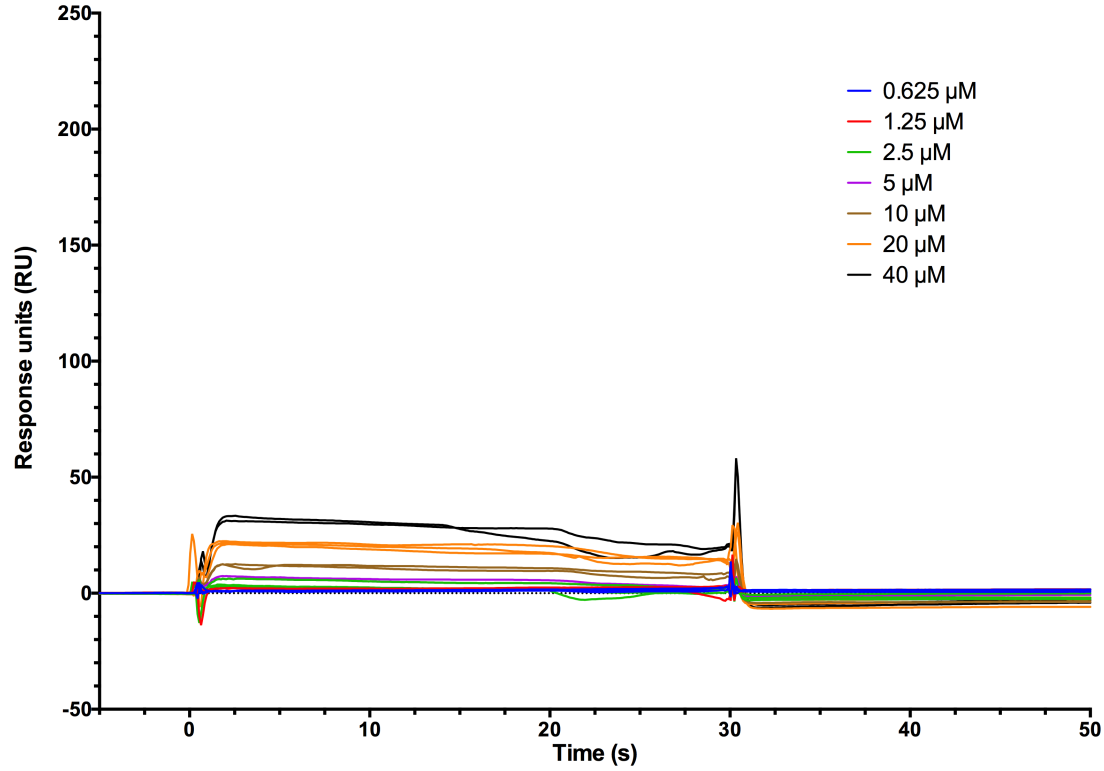

Figure 10.1.7. TBL1XR1 E171Q with MeCP2 285-313 peptide

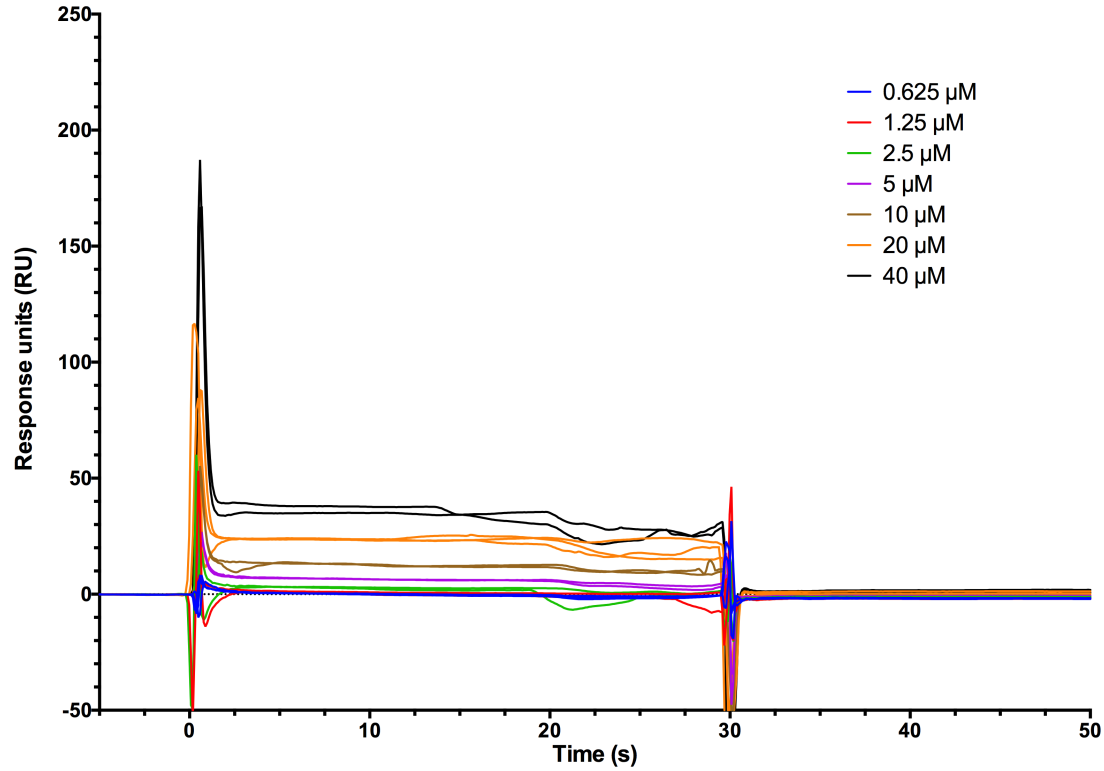

Figure 10.1.8. TBL1XR1 E171Q with MeCP2 285-309 peptide

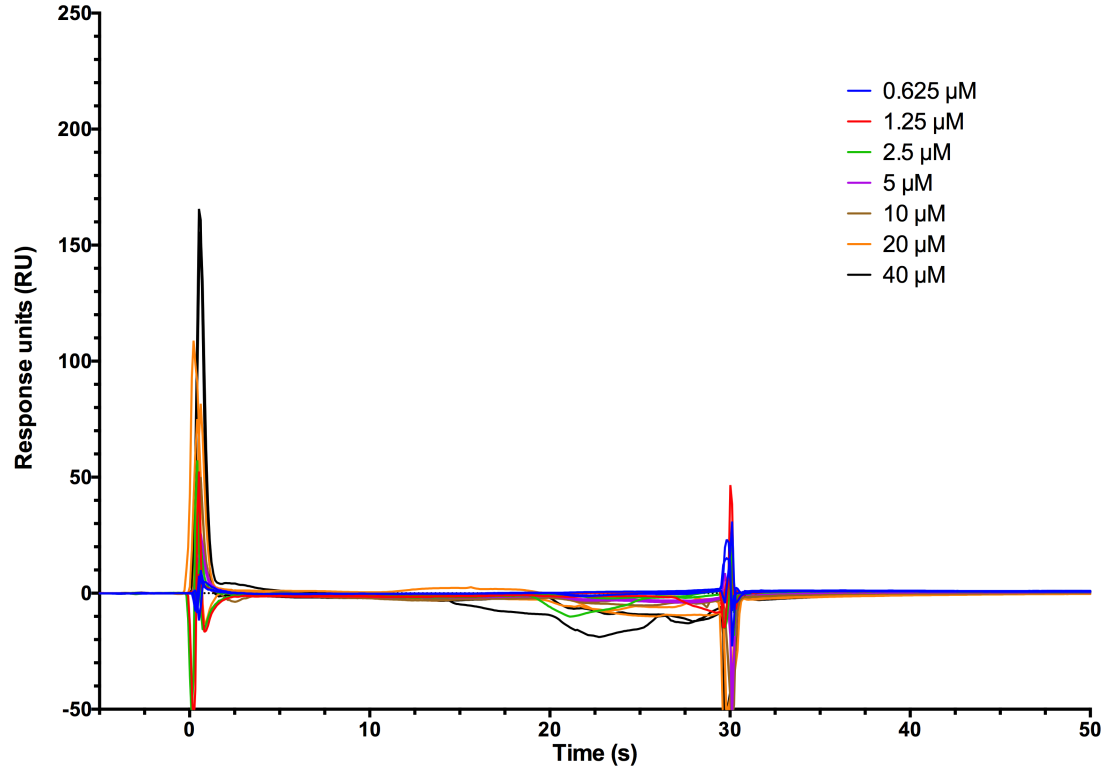

Figure 10.1.9. TBL1XR1 E171Q with MeCP2 285-313 R306C peptide

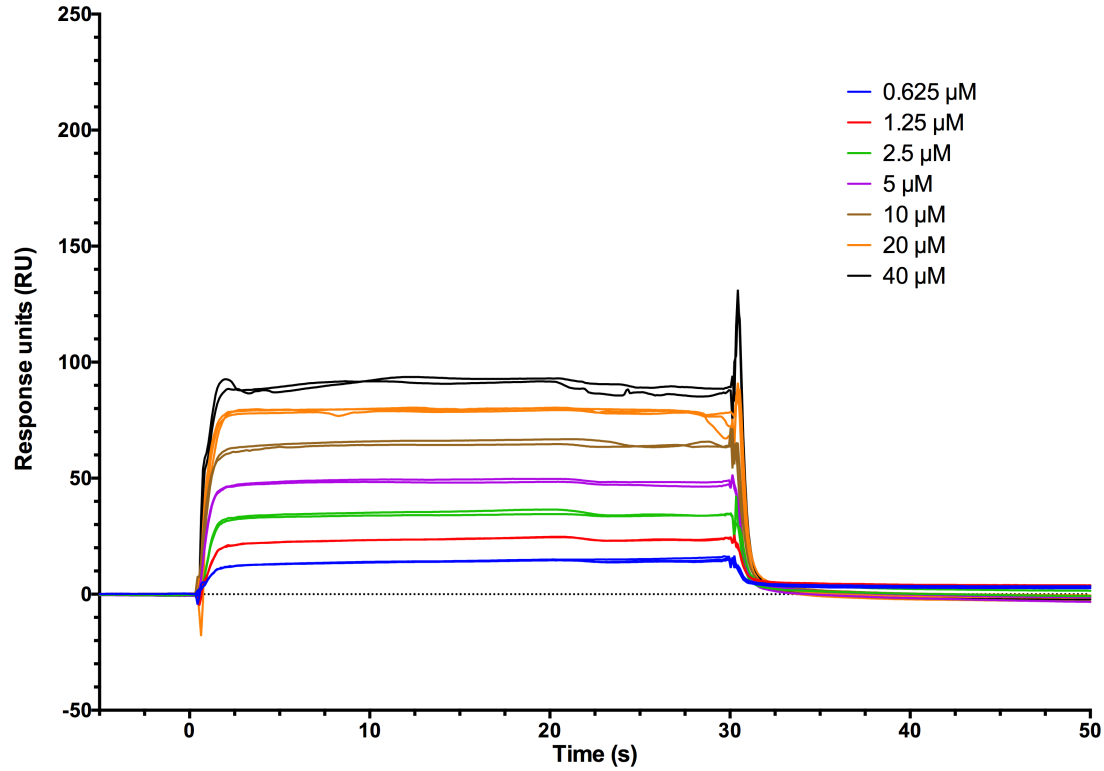

Figure 10.1.10. TBL1XR1 C214S with MeCP2 285-313 peptide

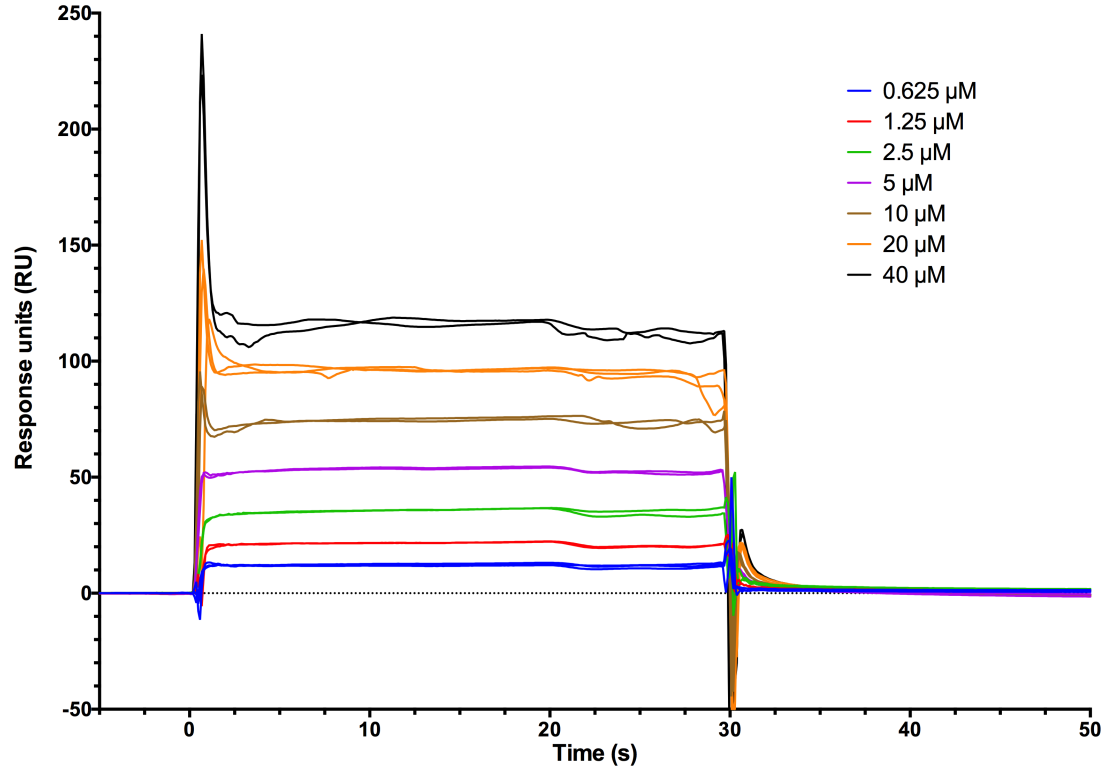

Figure 10.1.11. TBL1XR1 C214S with MeCP2 285-309 peptide

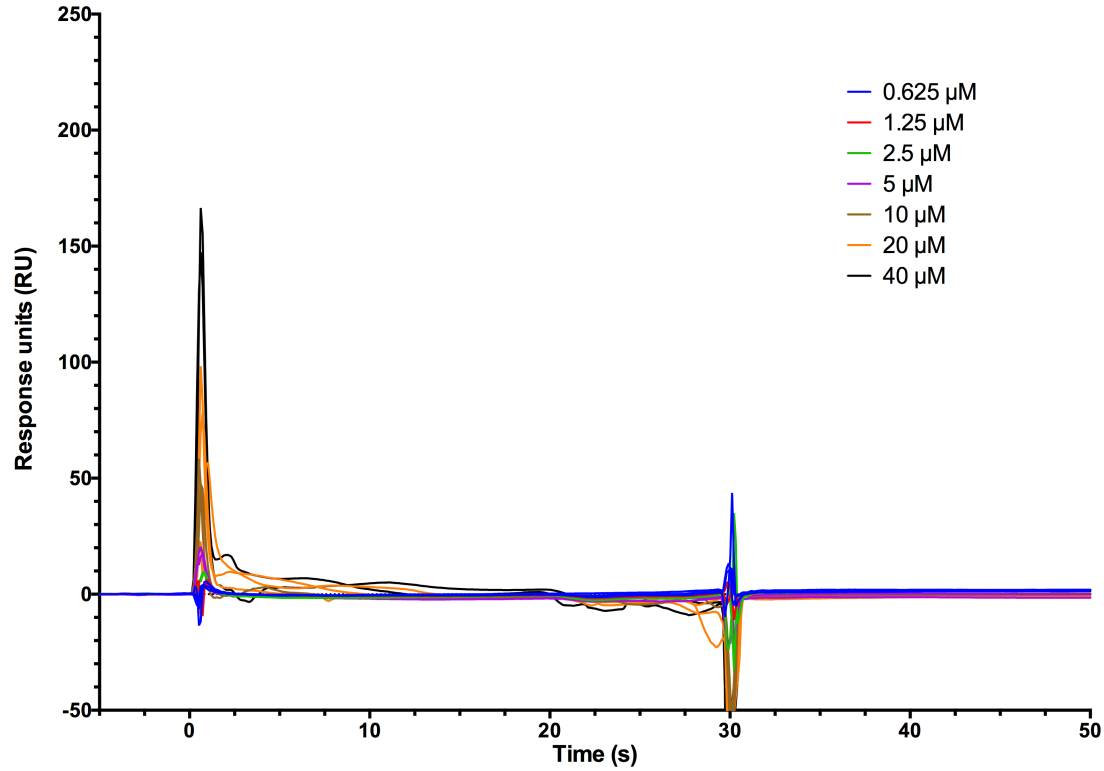

Figure 10.1.12. TBL1XR1 C214S with MeCP2 285-313 R306C peptide

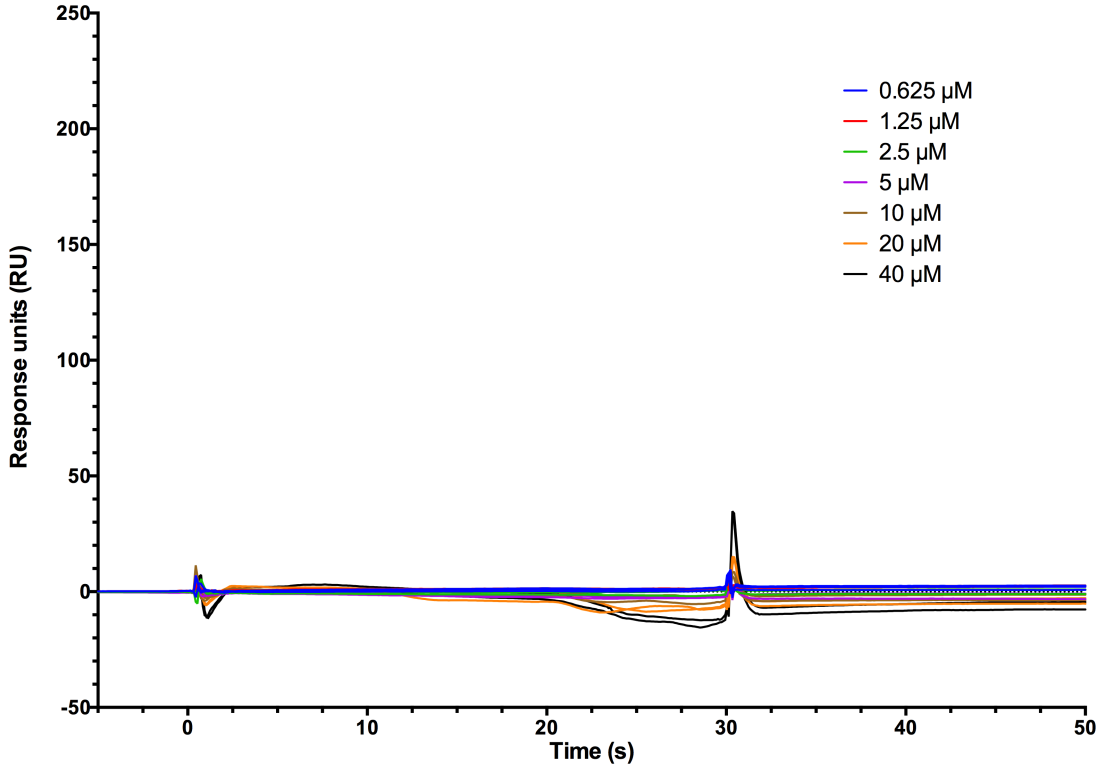

Figure 10.1.13. TBL1XR1 D313N with MeCP2 285-313 peptide

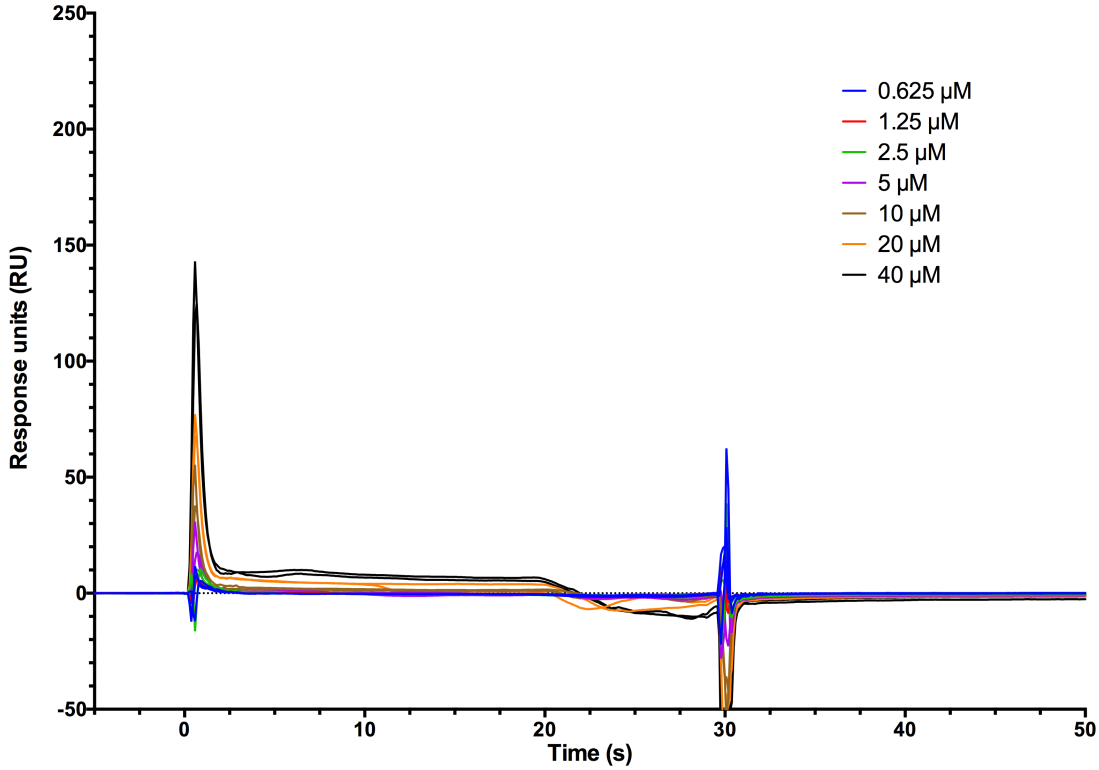

Figure 10.1.14. TBL1XR1 D313N with MeCP2 285-309 peptide

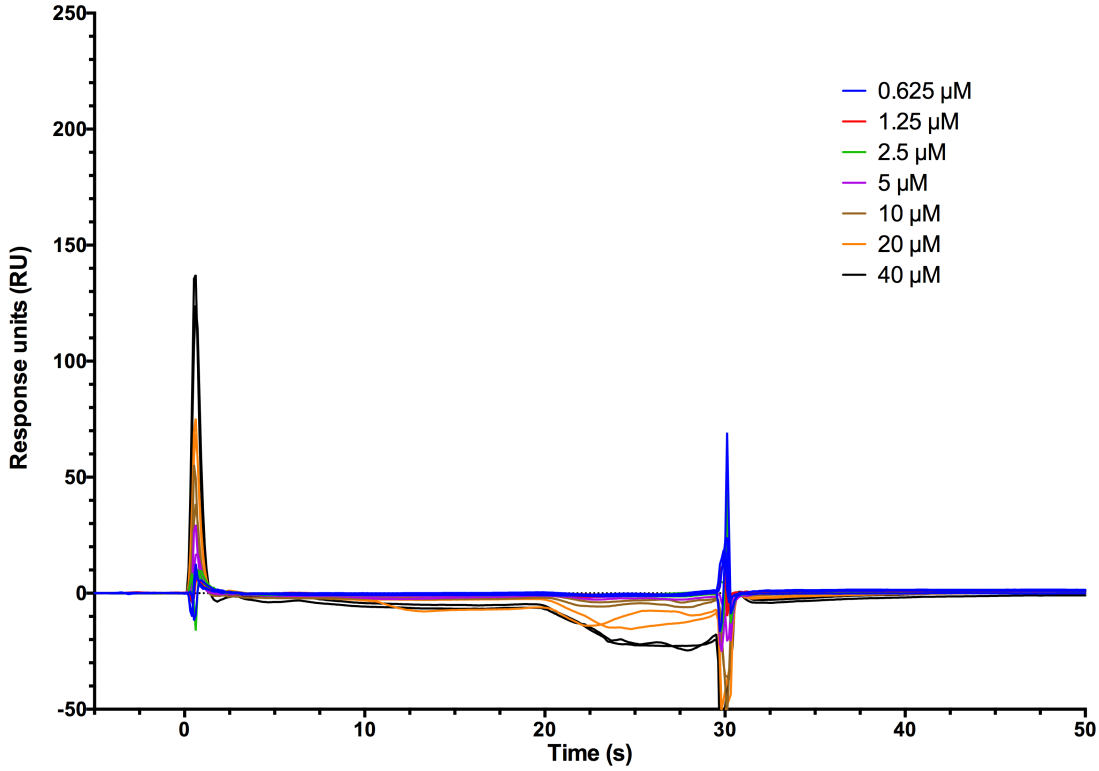

Figure 10.1.15. TBL1XR1 D313N with MeCP2 285-313 R306C peptide

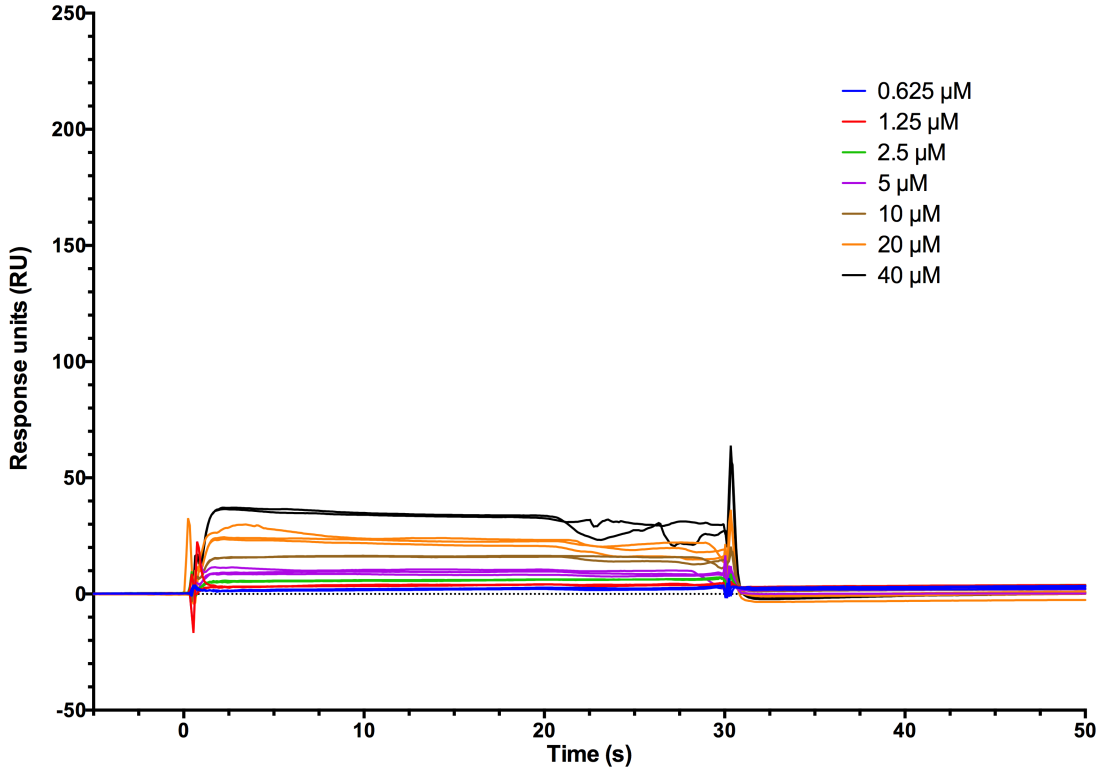

Figure 10.1.16. TBL1XR1 E351A with MeCP2 285-313 peptide

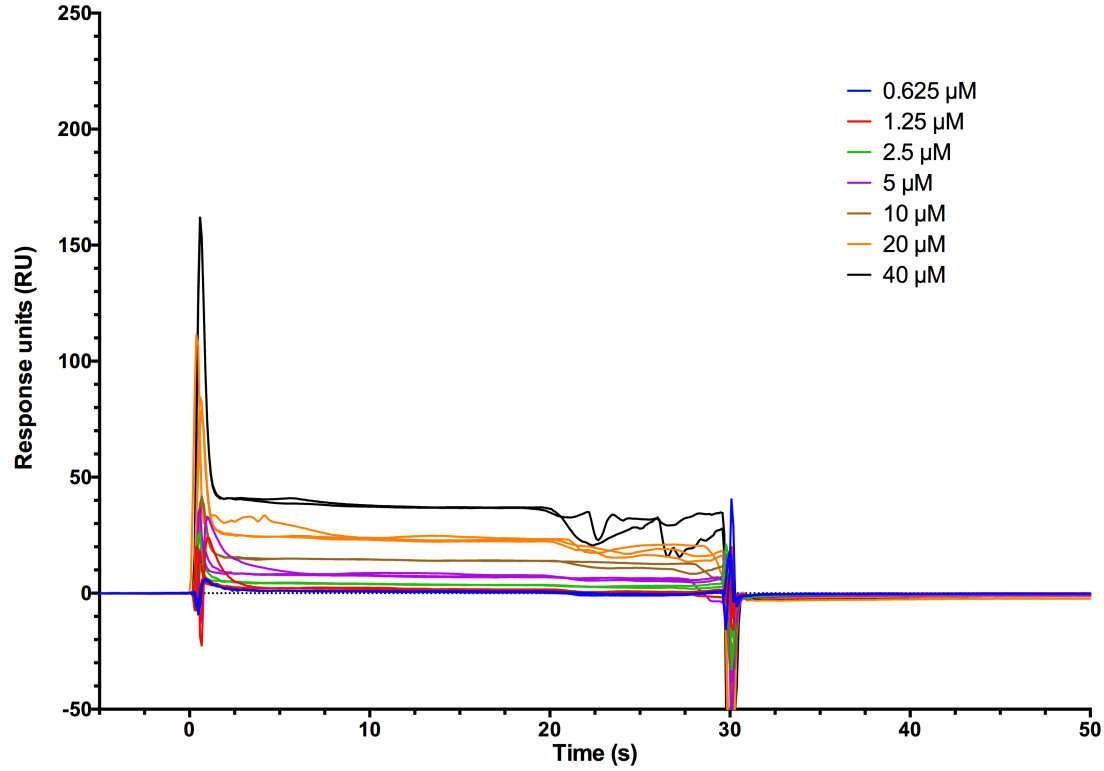

Figure 10.1.17. TBL1XR1 E351A with MeCP2 285-309 peptide

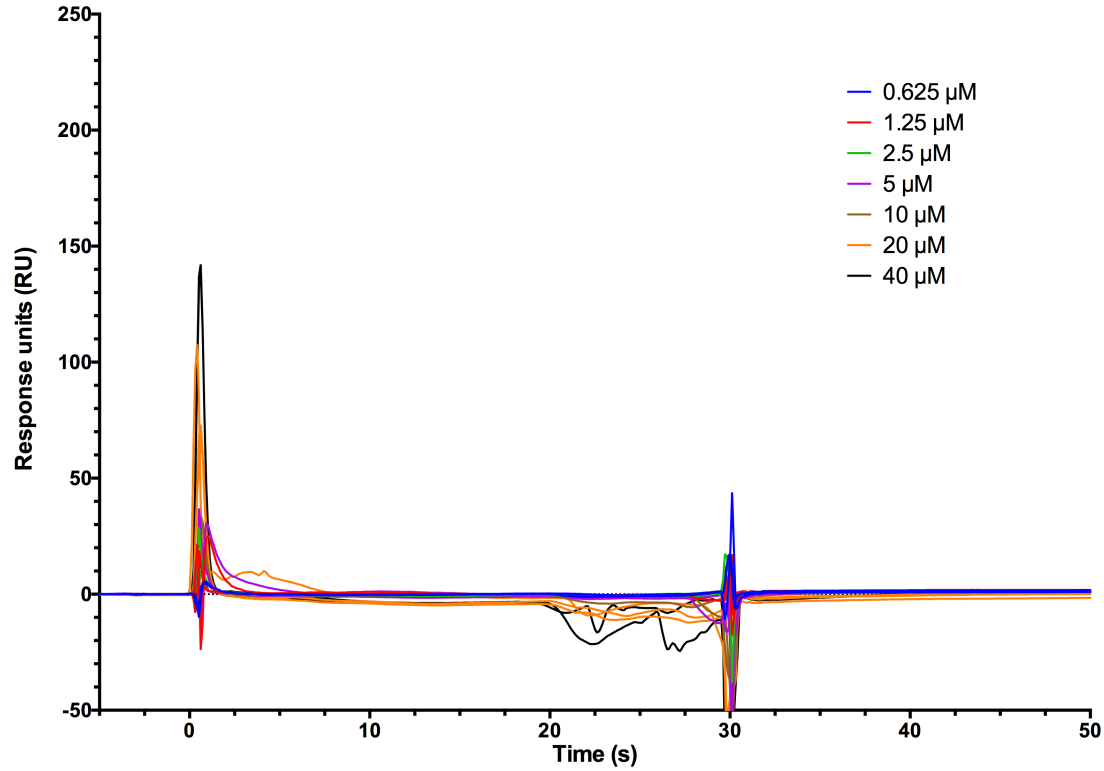

Figure 10.1.18. TBL1XR1 E351A with MeCP2 285-313 R306C peptide

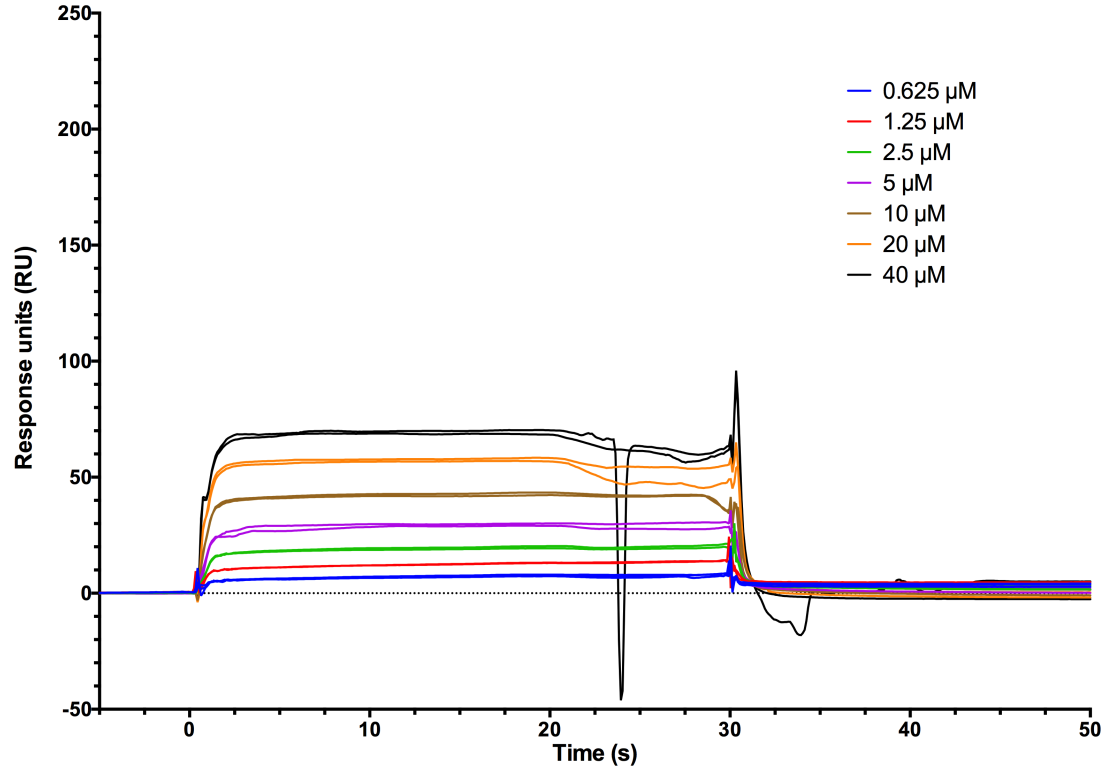

Figure 10.1.19. TBL1XR1 E351D with MeCP2 285-313 peptide

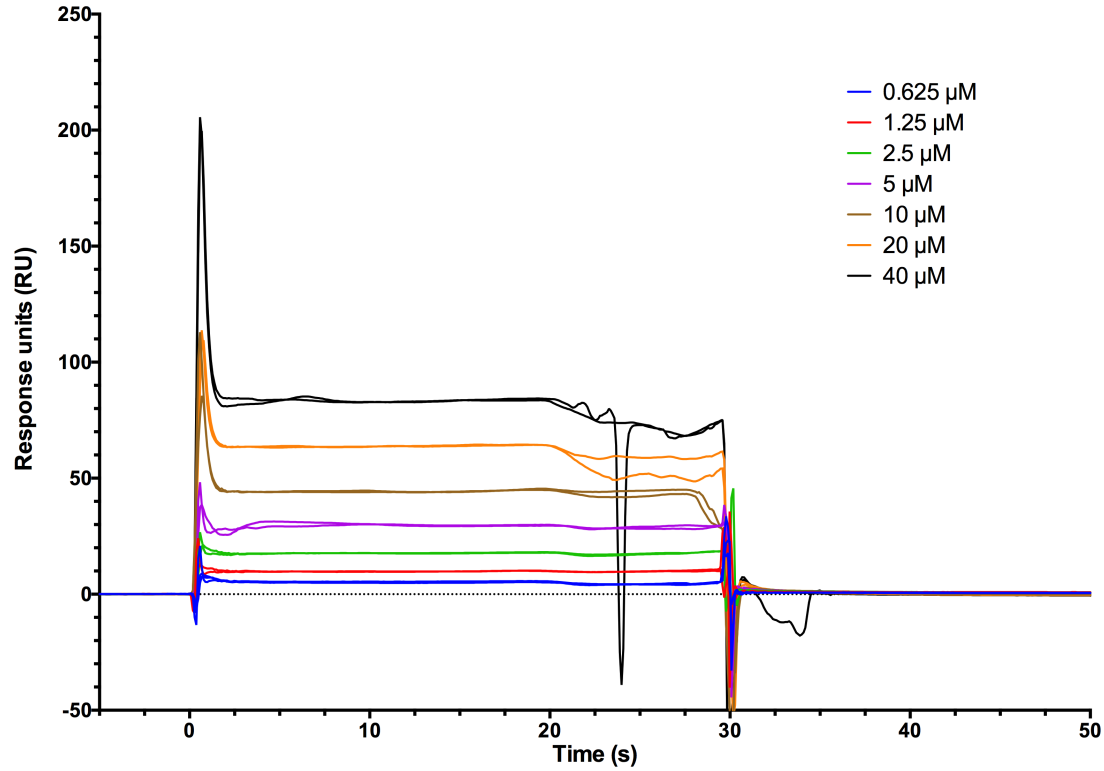

Figure 10.1.20. TBL1XR1 E351D with MeCP2 285-309 peptide

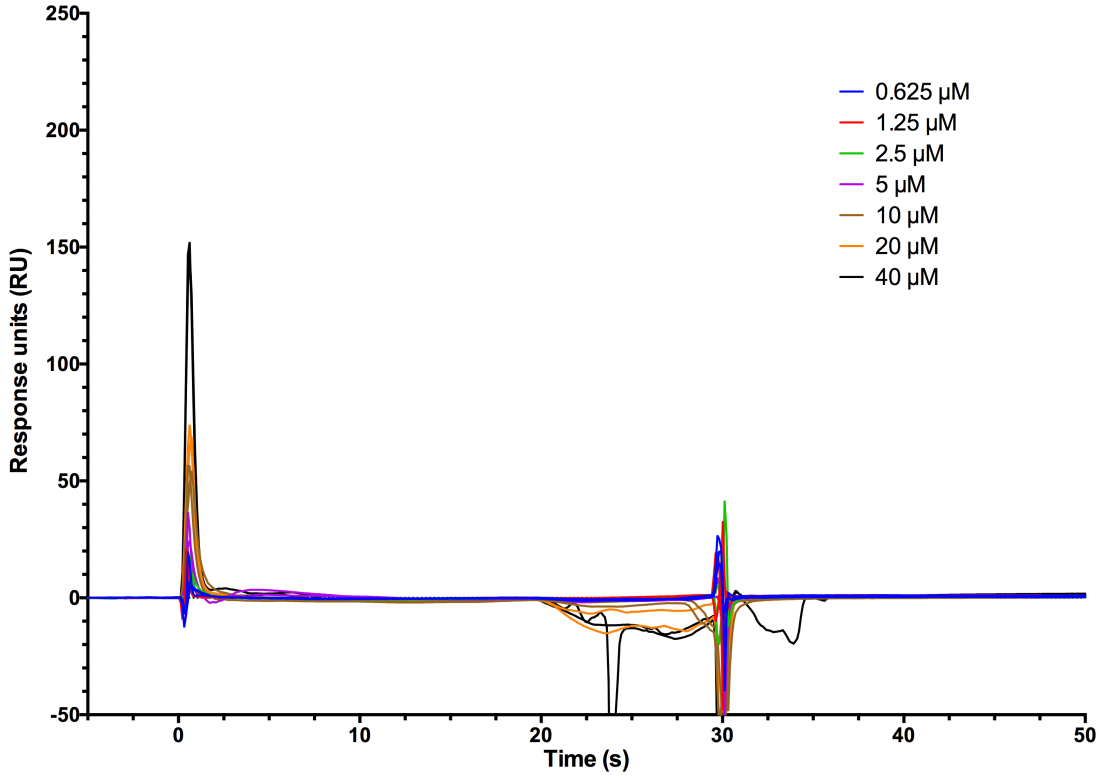

Figure 10.1.21. TBL1XR1 E351D with MeCP2 285-313 R306C peptide

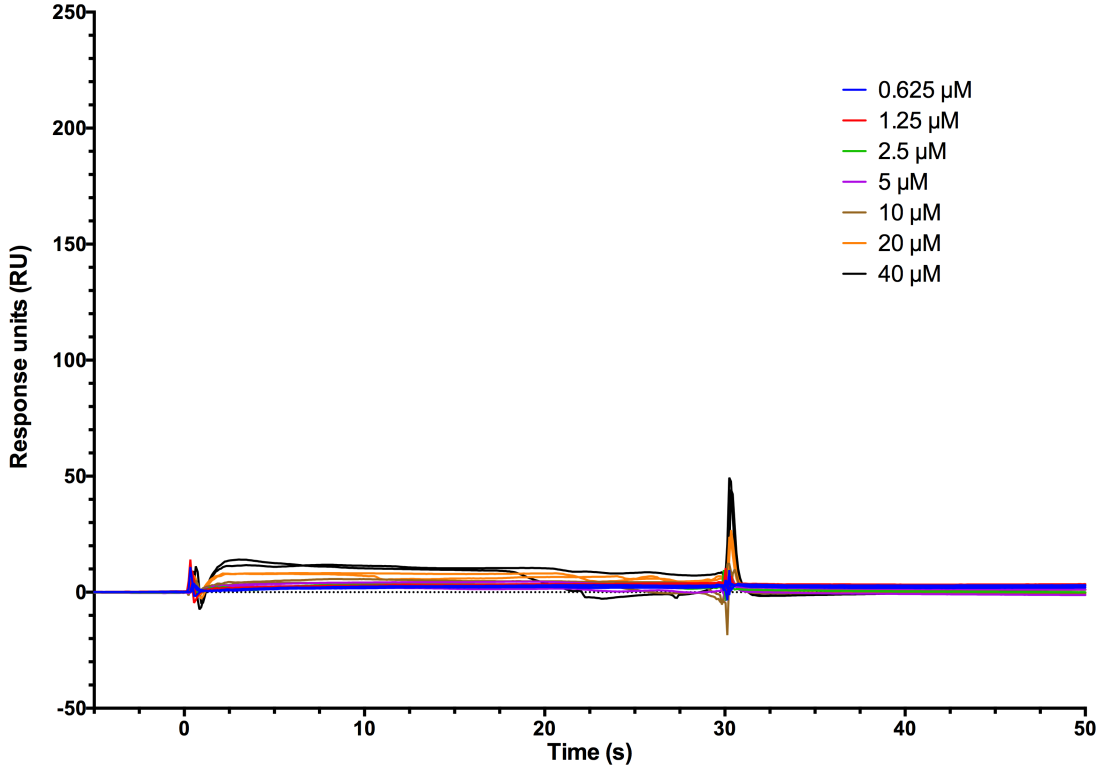

Figure 10.1.22. TBL1XR1 D369A with MeCP2 285-313 peptide

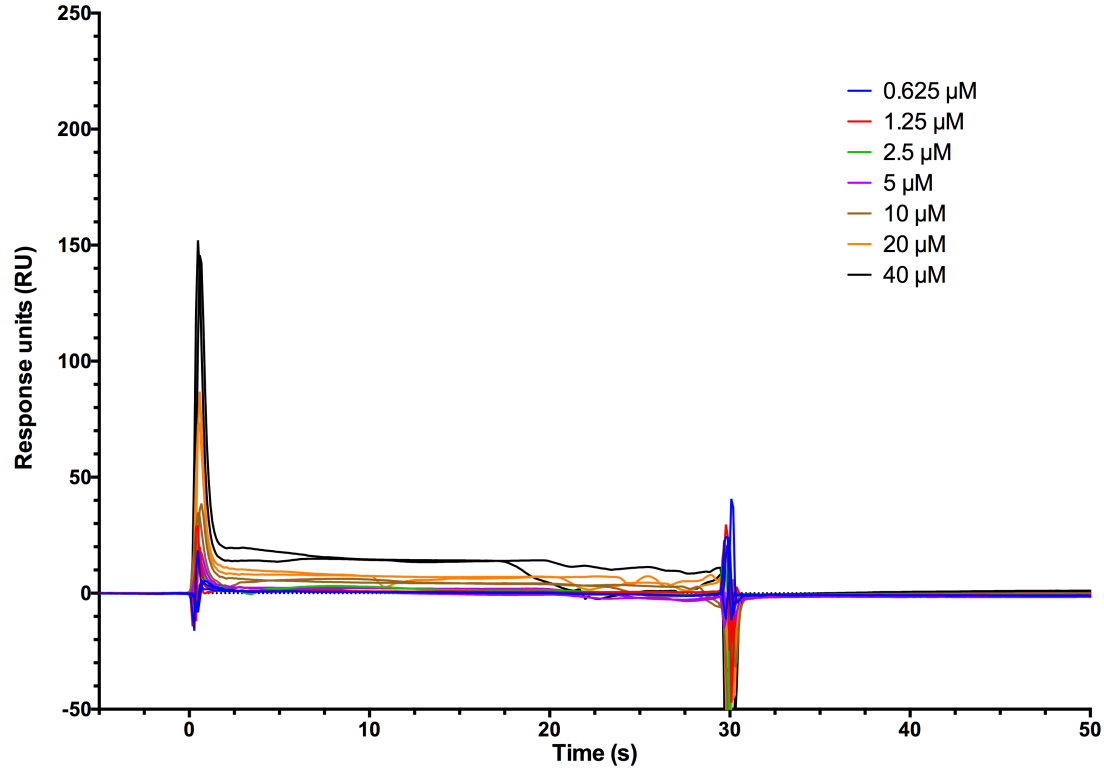

Figure 10.1.23. TBL1XR1 D369A with MeCP2 285-309 peptide

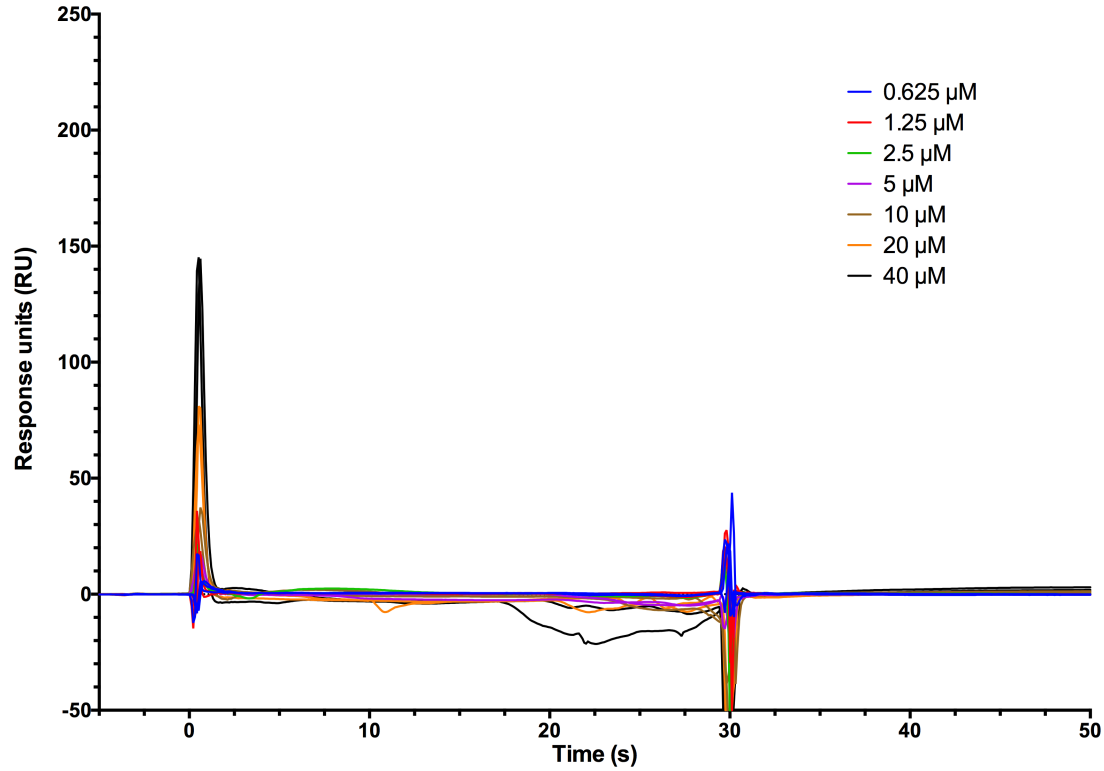

Figure 10.1.24. TBL1XR1 D369A with MeCP2 285-313 R306C peptide

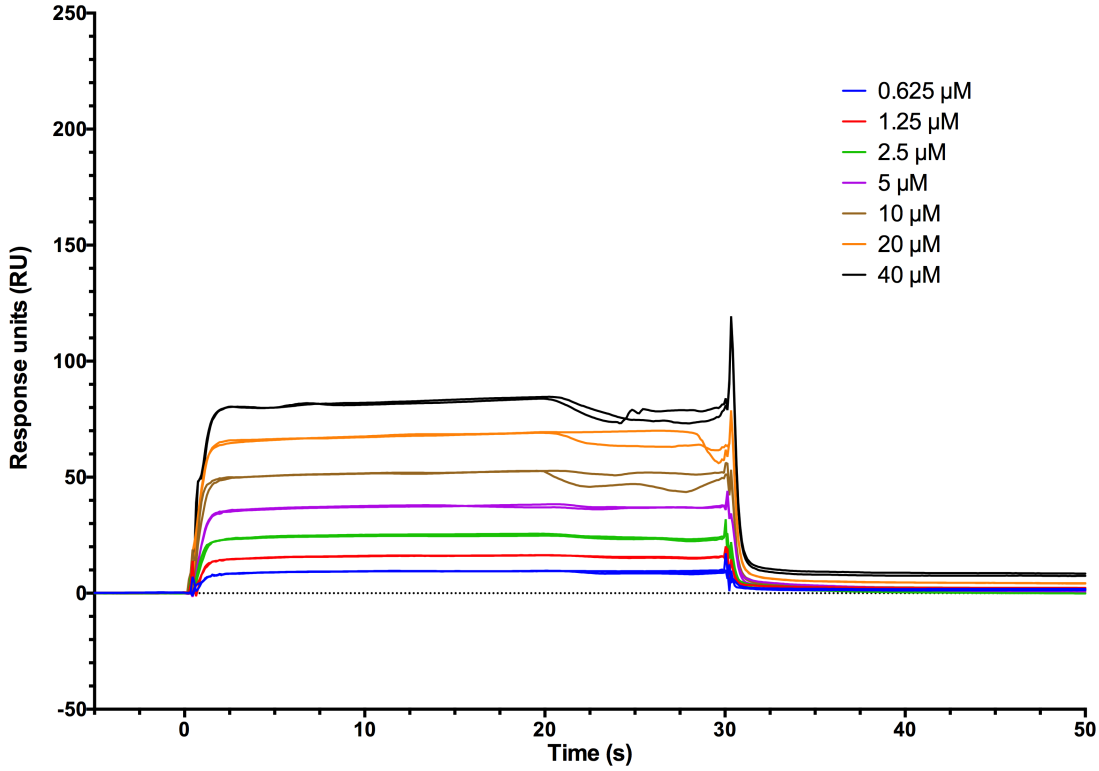

Figure 10.1.25. TBL1XR1 Y446F with MeCP2 285-313 peptide

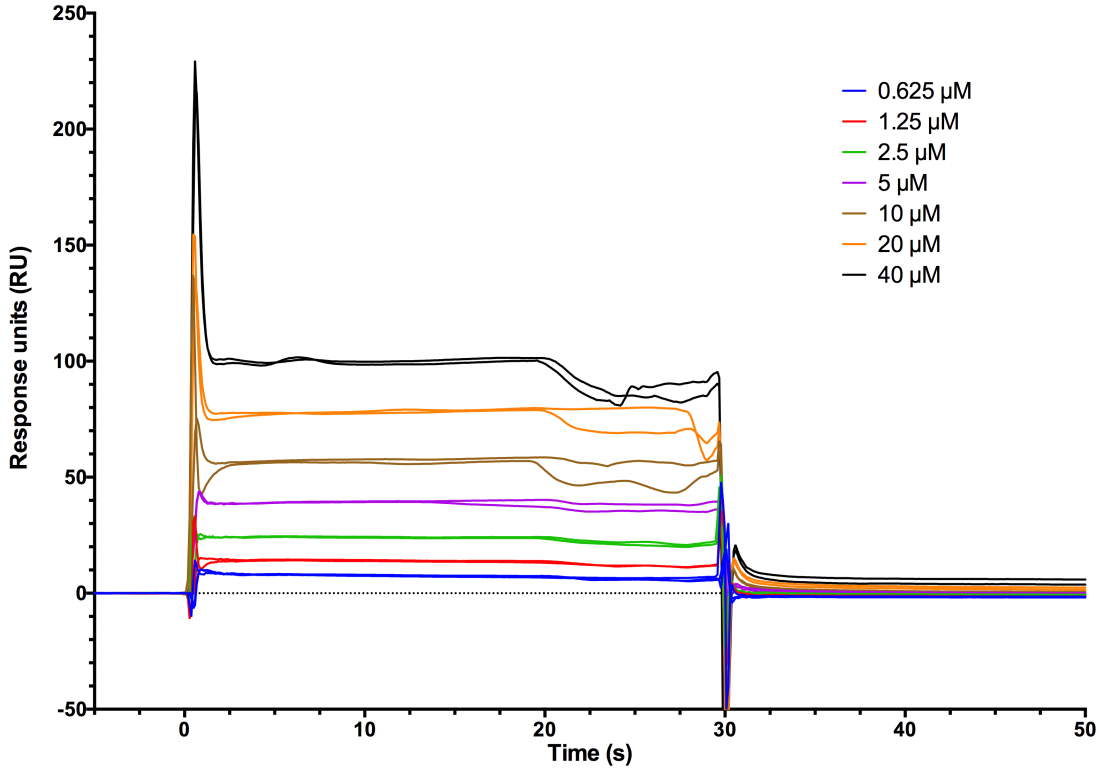

Figure 10.1.26. TBL1XR1 Y446F with MeCP2 285-309 peptide

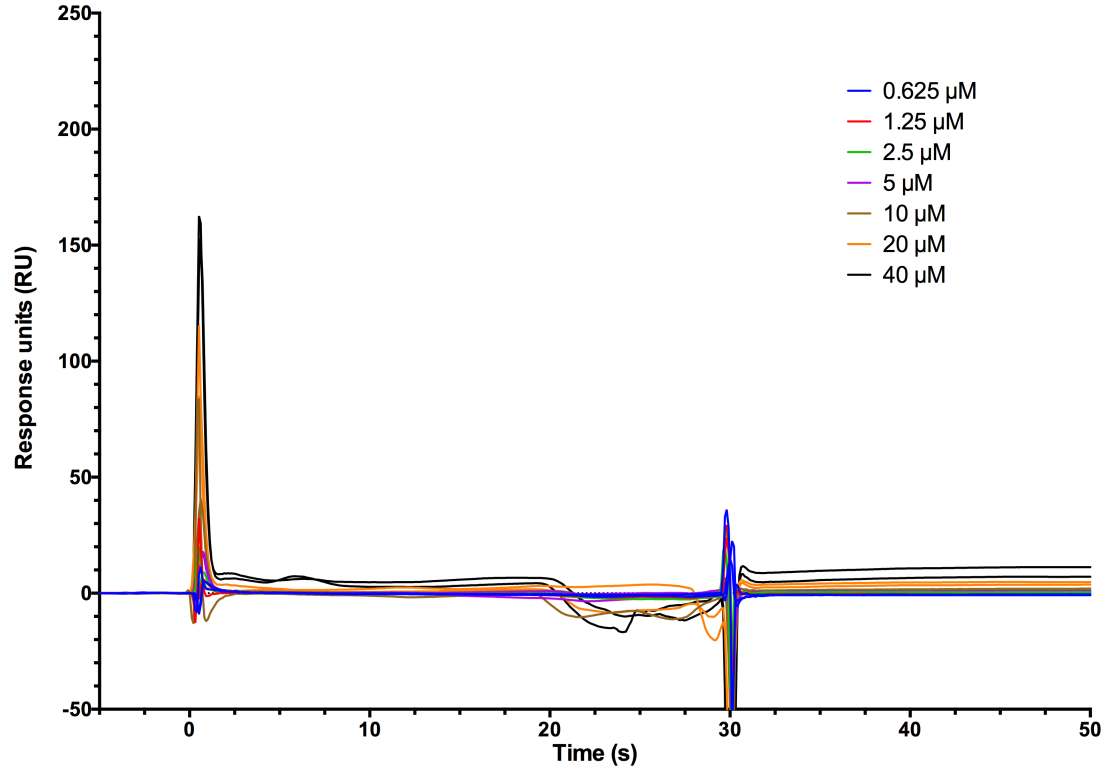

Figure 10.1.27. TBL1XR1 Y446F with MeCP2 285-313 R306C peptide